ECONOMIC STUDIES

DEPARTMENT OF ECONOMICS
SCHOOL OF BUSINESS, ECONOMICS AND LAW
UNIVERSITY OF GOTHENBURG
191

Economic Implications of Corporate Social Responsibility and Responsible Investments

Cristiana Mănescu

UNIVERSITY OF GOTHENBURG

ISBN 978-91-85169-53-5
ISSN 1651-4289 print
ISSN 1651-4297 online

Printed in Sweden,
Geson Hylte Tryck 2010

To my family,

Always near, no matter how far

Contents

Paper III: Stock returns in relation to environmental, social, and governance performance: Mispricing or compensation for risk?

Paper IV: Doing worse while doing less good

ACKNOWLEDGEMENTS

Realizing that I have now managed to publish my first book, I would like to pay homage and offer my gratitude to those who have given me inspiration, guidance, and support – I would never have been able to do it without you.

I would like to first thank my first supervisor, Cătălin Stărică, who offered me the amazing opportunity to pursue this path, tried to teach me the rigor of science, and always encouraged me to enjoy the journey just as much as reaching its end. To my second supervisor, Martin Holmén, goes all my gratitude for his key guidance throughout the final year, his instrumental contribution in helping me meet all the deadlines, and his timely comments and suggestions on the many revised versions of my papers. Many thanks to both of you! Equally much, I would like to thank MISTRA for generous financial support and for facilitating both inspiring and enjoyable meetings with top researchers in the area of responsible investment.

Several other professors and researchers have contributed to my current knowledge and skills. In doing so, they have helped me develop into the junior Economist I am today. Among them, in no particular order, I would like to kindly thank Evert Carlsson for always challenging my statements and arguments and for always speaking Swedish to me, Måns Söderbom for his excellent lectures in Econometrics and for always keeping his door and mind open to my many questions, Olof Johansson Stenman for his passionate Economics lectures, Lennart Flood for motivating us through his excellent Econometrics lectures, Stefan Sjögren and Mattias Hamberg for helping broaden my corporate finance and accounting research horizons, Frauke Skudelny for trusting my skills, inviting me to work with her at the European Central Bank, and for always being a source of inspiration and motivation, Lennart Hjalmarsson for being so supportive ever since my first month in Sweden and at Handelshögskolan, and Leonardo Becchetti, Hossein Asgharian, and Rob Bauer for their valuable comments and suggestions on this thesis.

For administrative support, and even sometimes lending an ear or a helpful hand to my screams of joy or despair, I am deeply grateful to Katarina Forsberg, Eva-Lena Neth Johansson, and Jeanette Saldjoughi. Equally much, I want to thank Rick Wicks and Debbie Axlid for significantly improving the quality of my written English.

My thanks also go to all my colleagues and friends at the Department of Economics and at the Department of Business Administration, and in particular to: Miyase Köksal-Ayhan, Andreea Mitruț, Corina Miller (Lunganu), Daniela Andrén, Amrish Patel, Alexander Herbertsson, Florin Maican, Aihua Xu, Jonas Gustafsson, Markus Rosenberg, Louise Holm, Kalle Erlandzon, Roger Wahlberg, Taylan Mavruk, Conny Overland, Hans Jeppsson, Georgios Foufas, and Mattias Sundén for sharing the ups and downs and the Economics talks, for the many laughs we had, races we ran, trips we made, and other moments we shared. Other friends, near and far, old and new, have constantly offered their support and showed interest in my work and experience as a

PhD student: Jan Stevens, Christina Olivecrona, Irina Nicula, Cristina Radu, Anca Apostoae (Robu), Radu Apostoae, Claudiu Komartin, Iulian Grosu, Leo Surugiu (Sandu), Magda Bodzioch, Stéphanie Giamporcaro, and Stacy Edgar – you will always be near to my heart!

My parents, Augustin and Maria, have always played an important part in my life. They are my heroes for successfully meeting – with grace, honesty, and hard work – all the challenges inherent in a society and economy in transition, for valuing education so highly, and for dedicating their entire life to their children (my wonderful brother, George, and me). I hope that one day I will make them as proud of me as I am of them. And my dear Constantin, thank you for making me feel special every day!

Frankfurt, October 5th 2010

ABSTRACT

Paper 1 (with Catalin Starica): This study conducts an in-depth analysis of the association between a unique ten-dimensional set of Corporate Social Responsibility (CSR) scores and firm profitability, as measured by Return on Assets (ROA). We find that non-linear (semi or non-parametric) regression methods bring important improvements in explaining profitability relative to a classical linear approach. While a number of CSR variables like corporate governance, talent attraction and codes of conduct might have some explanatory power, the CSR scores do not improve over the standard variables known to be associated with ROA.

Paper 2 (with Constantin Belu): This paper proposes a novel Corporate Social Responsibility (CSR) index based on a Data Envelopment Analysis (DEA) model. Acknowledging the argument that companies might favor those CSR dimensions that provide strategic competitive advantages, we argue that the index can capture companies' strategic approach to CSR. Furthermore, our findings reveal a neutral relationship between this strategic CSR index and economic performance as measured by ROA and Tobin's Q, when controlling for firm unobserved heterogeneity and past economic performance. By contrast, an equally-weighted index of the same CSR indicators is found to be negatively related with ROA, which reinforces our claim that this specific DEA-based index is a measure of strategic CSR.

Paper 3: Using detailed data on seven environmental, social, and governance (ESG) attributes for a long panel of large publicly-traded U.S. firms during July 1992-June 2008, only community relations were found to have had a positive effect on risk-adjusted stock returns, which effect was not compensation for risk but could be due to mispricing. Additionally, a changing effect of employee relations was found from positive during July 1992-June 2003 to negative during July 2003-June 2008. The positive effect could be due to mispricing, whereas there is some evidence that the negative effect was compensation for low non-sustainability risk. A weak negative effect of human-rights and product safety indicators on risk-adjusted stock returns in the more recent period was also found to be likely due to mispricing. The implications are that certain ESG attributes might be value relevant but they are not efficiently incorporated into stock prices.

Paper 4: This paper investigates how annual abnormal returns react to current and past rating revisions in corporate responsible behavior in a panel data spanning 16 years. I find that increases in less responsible behavior led to persistent negative abnormal returns, which were particularly strong for the area of corporate governance, and weaker for product safety and the environment. These results are robust to concerns of endogeneity, i.e., that the negative stock price movements would lead to an update in the areas of social responsibility concerns. In contrast, increases in already strong responsible behavior did not generate a systematic reaction in stock returns.

JEL: C14, C22, C23, C26, C52, C67, G12, G14, G30, M14

Keywords: Corporate Social Responsibility, Strategic CSR, Socially Responsible Investments, Sustainability, Firm Profitability, Stock Returns, Statistical Learning Techniques, Variable Selection, Smooth Splines, Regression Trees, Data Envelopment Analysis, Difference-GMM, Risk-Factor Test, Market Efficiency, Control Functions Approach.

INTRODUCTION

Corporate Social Responsibility (CSR) related practices are becoming part of our daily lives. Shops sell an increasingly broad range of products with some sort of green labeling, restaurants show an eco-friendly alternative menu, more and more corporations seek to brand themselves as socially or environmentally conscious, and increasingly more money under professional management is invested in firms that meet some kind of ethical or (more recently) sustainability criteria. According to Portney (2008), it is now "virtually impossible to open the business section of the *New York Times, the Wall Street Journal, The Economist,* or any business publication without seeing mention of measures taken by some company to become more 'socially responsible.'" As these examples indicate, the realm of CSR is very broad and there is still no strong consensus on a definition for it. This makes measurement problematic, implying direct serious consequences for empirical research.

Despite this newly acquired visibility, the interest in the issue dates long back. Both business and academic communities have paid attention to corporate responsibility towards society and nature, already since the beginning of the 20th century. Initially, two contrasting theories in management science laid the foundation for the CSR discussion, i.e., *the shareholder view* of the firm and the *stakeholder theory*. One of the most visible proponents of the shareholder theory has been Milton Friedman, who, in his famous *New York Times* article of 1970, argues that "the sole social responsibility of business is to increase profits," i.e., shareholder value. The stakeholder theory, presented by Freeman (1984), emphasizes that managers should meet not only the requirements of stockholders, i.e., owners of the firm, but also those of a variety of stakeholders (e.g., consumers, employees, suppliers, local communities), whose support is crucial for the existence of the firm.

At the same time, the empirical literature has tried to identify whether CSR is harmful or helpful for profitability, with often conflicting findings, especially with respect to certain aspects of CSR (e.g., the environmental dimension[1]). Therefore, more recently, a new theoretical perspective has been developed. It claims that only certain dimensions of CSR can improve the performance of a firm and these are all dimensions that bring a competitive advantage, given the business model and the industry in which the firm operates. Seen as part of the profit-

1 See Section 3 "Empirical Evidence" in Chapter 1.

maximization strategy of the firm, it is called *strategic CSR* (Baron, 2001; Porter and Kramer, 2006).[2]

The continuous growth in CSR adoption by companies has been accompanied by an almost equal growth in a specific type of investment based on certain values or CSR criteria, i.e., Responsible Investing (RI). Initially, RI represented the alignment of one's investments to one's values (therefore called "values-based" investment, Kinder, 2005), e.g., by screening out companies in controversial business areas (tobacco, weapons, alcohol etc.). However, more recently, responsible investment has been linked tighter to the CSR debate and witnessed an extension of investment criteria to incorporate a broader range of issues, i.e., environmental, social, and governance (ESG) issues, that are believed to inform, in non-financial terms, about future financial performance (Kinder, 2005). As with CSR, the exact definition of ESG is elusive. It generally incorporates investment criteria influenced by values-based judgment (values-based RI) and/or non-financial information that might have an impact on future firm performance (value-based RI. Given the diversity of investment practices that fall under RI concept, one can expect both identical and higher or lower return performance relative to conventional investing can be expected. In Section 2 of Chapter 3, I propose a complete set of hypotheses about ESG impact on investment returns, i.e., the no-effect scenario, the mispricing scenario, and the risk-factor scenario, based on the economic content of the ESG information as well as investors' access to it.

The aim of this thesis is to deepen our understanding of how measures of CSR (interchangeable with ESG) affect firm profitability as well as its risk-adjusted stock returns. I try to achieve these goals by pursuing several distinct avenues. Although I acknowledge throughout the whole thesis that CSR can be analyzed from different perspectives depending on the motivations underlying a firm's or an investor's engagement with CSR, I do not argue in favor of one or another of its forms. Instead, I simply analyze them in relation to economic performance. Overall, the thesis is empirical in nature. Moreover, each chapter makes use of a different modeling framework, suitable to the research question it focuses on. I will continue with a more detailed presentation of the perspectives taken and the results obtained, without emphasizing too much the specifics of each methodology applied.

---

2 See Section 2.1 in Chapter 2 for a detailed discussion on strategic CSR and its implications.

I analyze the corporate social performance - corporate financial performance link from two major perspectives, i.e., a corporate finance and an investments perspective. From the corporate finance perspective, I estimate the impact of various measures of CSR on firm profitability such as Return on Assets (ROA) and Tobin's Q. The main original contribution consists of evaluating a *non-linear impact* of the various CSR measures on ROA, and of proposing a *new aggregation method* for the various CSR dimensions that facilitates the capture of the strategic dimension of CSR. From the investments perspective, I evaluate the relationship between stock returns and ESG performance. The main contribution is two-fold. First, I test two different explanations for the relation of returns to levels of various ESG dimensions: *mispricing or compensation for (non) sustainability risk.* Second, I evaluate the *adjustment mechanism of returns to revisions in ESG performance*.

One of the main difficulties with empirical research on CSR data is the issue of integrating the many quantitative and qualitative[3] CSR variables into a meaningful picture of firm CSR standing[4]. Throughout the analysis, I purposely keep a general view on CSR by using multinational and/or multi-industry samples, by using disaggregate measures, whenever possible, or by proposing *meaningful* aggregate measures of CSR.

The investigation I conduct is based on ratings provided by two reputable rating agencies: Sustainable Asset Management (SAM) and Kinder, Lydenberg, and Domini (KLD) Research & Analytics. These agencies provide background research for two widely cited sustainability and social responsibility indexes, i.e., the Dow Jones Sustainability Indexes and the Domini Social 400 Index, respectively.[5] The CSR data they provide differs structurally in terms of how construction and purposes served. SAM builds its data on questionnaires, which are usually filled out by the firms and have a clear focus on capturing non-standard information that might provide guidance with respect to various risks to firms' operations. KLD, on the other hand, does all its CSR research in-house, and its indicators tend to reflect values, ethics, and socially responsible issues, though several indicators are more economic/financial risk oriented.[6] The main strength of KLD data is its time span (it has been compiled since 1991), while the specificity of SAM data consists of its more proprietary nature (we are one of the few researches that have analyzed

3 An in-depth overview of this issue is provided in Section 4.1, Chapter 1.
4 Section 2.1., Chapter 2 provides a lengthy discussion on aggregation problems.
5 Thus, both CSR data sources regard only publicly listed firms.
6 KLD even has a category of indicators called Controversial Business Issues, which rates companies based on whether their business involves alcohol, gambling tobacco, firearms, military, or nuclear power.

this data). There are, of course, many other CSR data providers, especially ones that have came into the business in the last ten years, and they all differ in terms in terms of methodology, focus and geographical coverage.

In the **first chapter** of the thesis, we conduct an exploratory inquiry on the relevance of SAM's ten-dimensional CSR dataset to explain firm profitability over a five-year period (2002-2006). Initiated at a time of abundant conflicting empirical results on the linear relation between single CSR dimensions and accounting-based measures of profitability, this study is set out to investigate whether non-linearities in the association between *disaggregate* CSR scores and ROA might provide a more truthful picture. Based on their proven performance and ease of interpretability, we chose semi- and non-parametric non-linear methods developed and used by the statistical learning community.

The main two findings of the paper are of a negative nature. First, the SAM CSR scores do not contribute to explaining ROA. One cannot explain firm's ROA any better by adding the SAM CSR scores to the typical set of variables that are commonly believed to explain variation in ROA (change in sales, retained earnings, etc.). As they do not worsen performance by too much, one can still try to determine, using the two non-linear methods, which CSR dimensions might be related to ROA. Our second finding is that only few of the SAM's CSR variables (corporate governance, talent attraction, and codes of conduct) show some relevance to explaining firms' ROA. Their impact on ROA is non-linear as it generally levels off for extreme values. Thus, ROA is (up to one standard deviation) higher than average for above-average values in all of these variables, while only below-average values in corporate governance seem to hurt ROA. These findings might suggest that a careful selection of CSR indicators must be performed if a manager intends to improve firm profitability through CSR practices. Another important finding of our analysis, unrelated to CSR, is that many of the non-CSR variables, usually assumed to be linearly related to profitability, also show a significant non-linear relation to ROA, especially in the tails of the distribution.

The **second chapter** tries to address the difficulty that CSR researchers and business analysts face when having to aggregate the multifaceted CSR indicators into a unique quantitative measure of CSR standing and it was motivated by the relatively recent theoretical literature on "strategic CSR," as outlined above. Thus, we propose a novel method for measuring the strategic dimension of CSR that can be applied to a variety of CSR indicators and does not require prior knowledge on how strategic CSR should be measured. The method is based on Data Envelopment Analysis, a versatile non-parametric tool that assigns an endogenously determined set of weights to the various CSR dimensions. The endogenous weights are sample dependent and, therefore give the benefit of a relative measure: the aggregate CSR index of a firm is measured relative to the CSR performance of the other firms. When relating this measure of strategic CSR built on SAM's CSR scores to economic performance, measured either by ROA or Tobin’s Q, and controlling for firm-specific effects, we find a neutral relationship. Yet, we do find a negative relationship between an equally-weighted index of the CSR dimensions and ROA.

This chapter's main contribution is two-fold. First, our analysis shows that aggregation matters greatly when it comes to measuring CSR performance. Second, we provide a ready-to-use method of evaluating a firm’s CSR performance relative to its peers that could be used by rating agencies or other institutions that need an aggregated measure of CSR.

In **chapter three,** I introduce the concept of a sustainability risk factor, determined by the ESG performance of a firm that could be priced in the market once enough information on ESG or sustainability becomes available. This chapter is motivated by recent evidence that investment strategies of buying high ESG performers and selling low ESG performers could provide substantial premiums even over multi-annual periods of time (Derwall et al. 2005; Kempf and Osthoff, 2007). From a methodological perspective, I implement an asset-pricing test that indicates whether an observed relation, estimated in an earlier step, between stock returns and a candidate ESG risk factor is due to compensation for risk or to mispricing. The findings provide only weak evidence that an employee-relations mimicking portfolio behaves as a risk-factor on the limited period 2003-2008, while all other positive effects of community relations during 1992-2008 and employee relations during 1992-2003 and the weak negative effects of product safety and environment during 2003-2008 on returns might have been due to mispricing. While the evidence that the stock return differential could not be attributed to an increased/decreased (non-) sustainability risk was conclusive, the mispricing hypothesis was neither confirmed nor rejected, thus requiring further research. This type of analysis should be replicated on an ESG dataset that better reflects sustainability and, thus, a more representative sustainability mimicking portfolio could be built. The main takeaway is that there appears to be a shift in how ESG is perceived in the market, presumably due to information availability and to awareness.

The **final chapter** shows that while stock returns are generally responsive to revisions in areas of negative responsible behavior (as measured by the KLD weaknesses), they are irresponsive to revisions in areas of positive social behavior (as measured by the KLD strengths). It is mainly revisions with respect to weaknesses in corporate governance, product safety, and environment that have a negative effect on returns. Moreover, this effect is causal as no indication was found, using a flexible control functions approach, that past negative returns would trigger a revision in the number of weaknesses. By aggregating strength and weakness indicators into a single ESG measure, as commonly done in earlier studies using KLD data (including my earlier chapter), one might be led to imply a uniform reaction in stock reaction to both sets of indicators, although this analysis clearly shows this is not the case. This study was motivated by the important role that availability of information on ESG plays for market valuations, as implied also in the previous chapter. The findings indicate that information related to negative social responsibility events is value-relevant for investments, while information related to positive events is not.

Implicațiile economice ale responsabilității sociale a companiilor şi ale investițiilor cu considerente sociale

Activitățile legate de responsabilitatea socială a companiilor au devenit o prezență constantă în viața de zi cu zi. În magazine se vând din ce în ce mai multe produse cu embleme ecologice, în restaurante se oferă meniuri cu produse ecologice şi tot mai multe companii fac eforturi ca să îşi promoveze preocuparea față de comunitatea în care operează sau de mediul înconjurător. Din ce în ce mai mulți bani sunt investiți în companii care îndeplinesc criterii etice sau, mai recent, sustenabile. Portney (2008) notează că este imposibil să deschizi secțiunea de afaceri a ziarelor fără ca să vezi mențiuni despre măsuri luate de companii cu scopul de a deveni „responsabile social". Aşa cum arată exemplele menționate, noțiunea de responsabilitate socială a companiilor (RSC) este foarte cuprinzătoare şi încă nu există un consens cu privire la o definiție adecvată. Acest fapt are consecințe serioase pentru cercetările empirice.

În ciuda vizibilității recent câştigate, interesul față de acest concept are o istorie îndelungată. Atât comunitatea academică cât şi cea de afaceri au avut în vedere aspectul responsabilității companiilor față de mediu şi de societate încă de la începutul secolului 20. Inițial, două teorii din management au aşezat fundațiile discuțiilor legate de RSC, *perspectiva acționarilor* şi, respectiv, *perspectiva deținătorilor de interese*. Unul dintre cei mai cunoscuți susținători ai teoriei *perspectivei acționarilor* a fost Milton Friedman, care, într-un articol faimos din New York Times din 1970, a susținut că „singura responsabilitate a companiilor este să îşi crească profitul". Teoria *perspectivei deținătorilor de interese*, propusă de Freeman (1984), subliniază că managerii ar trebui să urmărească nu numai cerințele acționarilor, respectiv ale proprietarilor companiei, ci şi pe cele ale unei game largi de deținători de interese în companie (consumatori, angajați, furnizori, comunități locale), al căror suport este crucial pentru funcționarea companiei.

În acelaşi timp, literatura de cercetare empirică a încercat să identifice dacă RSC afectează negativ sau pozitiv profitabilitatea, iar rezultatele evidențiate sunt de multe ori contradictorii, în special în legătură cu anumite aspecte ale RSC (de ex. dimensiunea legată de mediu). Recent s-a dezvoltat o nouă abordare teoretică care susține că doar anumite aspecte ale RSC pot să îmbunătățească performanța firmelor. Acestea sunt dimensiunile RSC care asigură avantaje competitive, în funcție de modelul de afacere şi de industria în care compania operează. Văzută ca parte a unei strategii de maximizare a profitului, această abordare a fost denumită RSC strategică (Baron 2001; Porter şi Kramer, 2006).

Adoptarea în ritm crescător a principiilor RSC de către companii a fost acompaniată de o creştere aproape echivalentă a investițiilor în titluri de acțiuni, ale căror criterii de selecție se bazează şi pe anumite valori morale sau pe îndeplinirea unor criterii legate de RSC, aşa numitele „investiții responsabile" (IR). Inițial, conceptul de investiție responsabilă reprezenta alinierea strategiei de investiții cu valorile morale ale investitorului, de exemplu, evitarea investirii în companii care desfăşoară activități în domenii controversate (tutun, arme, alcool, etc.). Recent, IR au fost alăturate mai strâns dezbaterii din jurul RSC, şi criteriile de investiţie s-au extins prin încorporarea unei game largi de aspecte legate de mediu, societate şi guvernanţă corporatistă (MSG), aspecte care se crede că pot conţine informaţii ne-financiare despre evoluţia financiară viitoare a companiei în cauză. Ca şi în cazul RSC, nu există o definiţie clară cu privire la MSG. În general, conceptul de investiţie responsabilă implică criterii investiţionale bazate pe valori morale şi/sau informaţii ne-financiare care pot avea un impact asupra performanţei viitoare a firmelor.

Dată fiind diversitatea de principii de investiţii care pot fi catalogate drept IR, se poate aştepta atât un randament investiţional identic, cât şi unul mai mare, respectiv mai mic, în comparaţie cu investiţiile tradiţionale. În Secţiunea 2 a Capitolului 3, propunem un set complet de ipoteze vizând impactul criteriilor MSG asupra randamentelor titlurilor de acţiuni, respectiv scenariul lipsei de efecte, scenariul evaluării greşite şi scenariul factorilor de risc, bazate pe conţinutul economic al informaţiilor legate de MSG precum şi pe accesul investitorilor la aceste informaţii.

Scopul acestei lucrări este să aprofundeze cunoştinţele noastre despre modul cum implementarea măsurilor legate de RSC (respectiv de MSG) afectează profitabilitatea firmelor, precum şi randamentul acţiunilor acestora, ajustat cu riscul aferent. Încerc să ating aceste obiective urmând câteva direcţii de cercetare. Deşi menţionez explicit în teza de faţă că RSC poate fi analizat din perspective diferite, care sunt determinate de motivaţia firmei sau investitorului de a adopta RSC, nu susţin o anumită perspectivă în particular, ci doar analizez angajamentul faţă de RSC în raport cu performanţa economică. În mare, teza de faţă are o abordare empirică. În plus, fiecare capitol conţine o metodologie diferită, adecvată aspectului cercetat. În cele ce urmează, voi prezenta pe scurt contribuţia fiecărei lucrări, evidenţiind rezultatele fără a insista asupra metodelor econometrice folosite.

Contribuţia principală **în prima lucrare** constă în folosirea de *metode neliniare* pentru estimarea efectelor avute de măsurile de RSC asupra profitabilităţii firmei. Un prim rezultat indică faptul că, per ansamblu, masurile de RSC folosite nu contribuie la o îmbunătăţire a calităţii predicţiei profitabilităţii firmei, ceea ce nu este surprinzător. Un al doilea rezultat sugerează că, dintre cei zece indicatori de RSC folosiţi, numai cei cu privire la guvernanţa corporatistă şi atragerea de personal calificat superior au un efect pozitiv dar neliniar asupra profitabilităţii.

În **a doua lucrare** se discută pe larg problemele conceptuale şi metodologice întâmpinate la agregarea într-o măsură unică a indicatorilor de RSC de natură cantitativă dar şi calitativă, atât de diferiţi între ei, care să şi fie reprezentativă pentru firme din industrii şi/sau regiuni diferite. Soluţia propusă constă în utilizarea unei noi *metode de agregare prin care* diferitele realizări de RCS ale unei companii sunt evaluate nu în mod absolut, ci *relativ* la cele ale altor companii cu un obiect similar de activitate, sau din aceeaşi regiune. Metoda de agregare se bazează pe tehnici de optimizare liniară şi este folosită pe scară largă în evaluarea eficienţei economice a firmelor. Se mai arată în lucrare, că un indicator agregat astfel construit evidenţiază şi caracterul strategic al iniţiativelor de RSC, după cum a fost prezentat mai sus. Analiza empirică a efectului acestui indicator nou construit asupra profitabilității relevă o relație neutra, subliniind astfel că inițiativele de RSC strategic sunt luate numai în măsura în care nu reduc profitabilitatea.

În **a treia lucrare** implementez un test empiric care să indice cauzele pentru care se observă un impact al diferitelor măsuri de MSG asupra randamentului titlurilor de acțiuni, şi anume dacă se verifică scenariul evaluării greşite sau cel al factorilor de risc. Pe o piață de capital care funcționează transparent şi eficient, prețurile acțiunilor reflectă aproape instantaneu toate informațiile publice referitoare la compania în cauză, astfel încât tranzacționarea pe baza exclusivă a astfel de informații nu poate conduce la obținerea un randament superior. Potrivit primului scenariu, randamentele titlurilor de acțiuni ale companiilor ce implementează criterii de MSG nu vor fi diferite de cele ale titlurilor obişnuite[7], dacă informația existentă cu privire la aceste criterii fie nu are un conținut economic relevant pentru companie, fie are un conținut economic relevant ce este însă complet reflectat în prețul acțiunii. Cel de-al doilea scenariu, al evaluării greşite, presupune că acest tip de informație are efecte economice, însă, deoarece investitorii nu au acces la ea (fie pentru că nu este făcută publică, fie pentru că este ignorată), aceasta nu poate fi corect încorporată în prețurile acțiunilor, i.e. prețul acțiunilor este evaluat greşit. Prin urmare, mai devreme sau mai târziu apar ”surprize” în e.g. profitabilitatea firmei, datorate criteriilor MSG, cu efect direct asupra randamentului acțiunilor, care poate fi astfel mai mare sau mai mic decât cel al unei firme obişnuite. Cel de-al treilea scenariu presupune că investitorii au acces perfect la astfel de informație (care este şi relevantă economic). Mai mult decât atât, investitorii o asociază cu un anume tip de risc ce poate afecta performanța financiară a companiei. Astfel, randamentele companiilor MSG pot fi diferite de cele ale companiilor obişnuite ca urmare a prezenței unui factor de risc financiar suplimentar. Rezultatele confirmă faptul că randamentele acțiunilor nu reacționează uniform la toate criteriile de MSG. Astfel, se arată că numai într-o anumită situație, şi anume în cazul criteriului legat de relația cu angajații, şi numai în perioada recentă 2003-2008, investitorii asociază un risc mai mic firmelor ce au relații bune cu angajații, şi de aceea acestea au un randament mai mic al titlurilor de acțiuni. În câteva alte situații, în legătură cu criteriile de mediu sau de siguranță a produselor, se confirmă scenariul evaluării greşite, însa în legătură cu cele mai multe criterii MSG nu se remarcă un randament diferențiat.

În **cea de-a patra lucrare**, analizez atent mecanismul prin care randamentele acțiunilor reacționează la informații noi despre nivelul de responsabilitate socială a firmelor, măsurată diferențiat de-a lungul unor dimensiuni negative, ce reflectă aspecte mai degrabă de iresponsabilitate sociala (de ex., atragerea de amenzi legate de poluarea mediului înconjurător), şi a unor dimensiuni pozitive, reflectând aspecte de excelență în responsabilitate sociala (de ex., guvernanța corporatistă transparentă). Rezultatele arată că înrăutățirea unei performanțe de RSC deja negative, în special în privința guvernanței corporatiste, a siguranței produselor şi a mediului, atrage o penalizare prin scăderea randamentului titlurilor de acțiuni, pe când îmbunătățirea în direcții considerate deja pozitive nu atrage nici o reacție.

[7] Prin companii obişnuite se întelege acele companii care nu implementează criterii de MSG.

References:

Baron, D. P., 2001, "Private politics, corporate social responsibility, and integrated strategy," *Journal of Economics & Management Strategy*, 10(1), 7-45.

Derwall J, Gunster N, Bauer R, Koedijk K., 2005, "The ecoefficiency premium puzzle," *Financial Analysts Journal* 61: 61–63.

Freeman, R.E., 1984, "Strategic Management: A Stakeholder perspective", *Englewood Cliffs,* NJ: Prentice Hall.

Friedman, M. 1970, "The social responsibility of business is to increase its profits," *New York Times Magazine*, September ,13.

Kempf A, Osthoff P., 2007, "The effect of socially responsible investing on portfolio performance," *European Financial Management* 13: 908–922.

Kinder P. 2005. Socially responsible investing: an evolving concept in a changing world Available at http:\\www:kld:com=resources=papers=SRIevolving050901:pdf: [Accessed 19.04.2010].

Porter, M. E. and M. R. Kramer, 2006, "Strategy and Society: The Link Between Competitive Advantage and Corporate Social Responsibility," *Harvard Business Review*, December, 78-92.

Portney, P., 2008, "The (Not So) New Corporate Social responsibility: An empirical perspective," Review of Environmental Economics and Policy, 2 (2), 261-275.

Paper I

Do Corporate Social Responsibility scores explain firm profitability? A case study on the publishers of the Dow Jones Sustainability Indexes[a][b]

Cristiana Mănescu[c] and Cătălin Stărică[d]

Abstract

This study conducts an in-depth analysis of the association between a unique ten-dimensional set of Corporate Social Responsibility (CSR) scores and firm profitability, as measured by Return on Assets (ROA). We find that non-linear (semi or non-parametric) regression methods bring important improvements in explaining profitability relative to a classical linear approach. While a number of CSR variables like corporate governance, talent attraction and codes of conduct might have some explanatory power, the CSR scores do not improve over the standard variables known to be associated with ROA.

JEL classification: C14, C52, G30, M14

Keywords: Corporate Social Responsibility, Firm Profitability, Statistical Learning Techniques, Variable Selection, Smooth Splines, Regression Trees

---

[a] *This research was conducted as part of the research programme Behavioral Impediments to Sustainable Investment, funded by the Swedish Foundation for Strategic Environmental Research, MISTRA.*

[b] **Acknowledgements:** We thank Carl-Johan Francke and Christophe Churet of Sustainable Asset Management (SAM) for providing clean data and for a careful and prompt reading of the manuscript, Evert Carlsson for his instrumental help in getting the data and for many fruitful discussions on it, Leonardo Becchetti and Hossein Asgharian for insightful comments and suggestions, and T. Hothorn for promptly answering all our R technical questions. We are also grateful for helpful suggestions and comments from participants at Mistra seminars at the University of Gothenburg, University of Umea, and Maastricht University, as well as at the Meeting of the Swiss Society of Economics and Statistics, St. Gallen, and participants at the Workshop on Quantitative Finance, Rome, the SEABUS Summer Academy, Berlin, the 1st Doctoral meeting of Montpellier, and the 16th annual EAERE conference. The views expressed in this paper are those of the authors and do not necessarily represent those of the SAM Group. All the corporate indicators in the present study have been used at the sole discretion of the authors of this article based on original data provided by the SAM Group. Any possible error in the interpretation or manipulation of such data remains the sole responsibility of the authors.

[c] Department of Economics/Centre for Finance, School of Economics, Business Administration and Law, Gothenburg, Sweden. Email: cristiana.manescu@cff.gu.se. Tel: +46 (0) 317864412

[d] Department of Economics and Department of Mathematical Statistics, at University of Gothenburg and Chalmers University of Technology, Gothenburg, Sweden. Institute of Statistics at University of Neuchatel, Neuchatel, Switzerland. Email: catalin.starica@unine.ch

1. Introduction

In its stronger form, the concept of Corporate Social Responsibility (CSR) asserts that corporations have an obligation to consider the interests of customers, employees, shareholders, and communities, as well as the ecological "footprint" in all aspects of their operations, which extends beyond their statutory obligation to comply with legislation. Initially understood as pure ethical or environmental concerns, CSR has developed into a complex concept spanning over three main directions, i.e., environmental, social, and governance issues (UNEPFI, 2007). One of the fundamental questions still to be answered concerns the effect of corporate socially responsible behavior on profitability. The answers one finds in the specialized literature are numerous and often contradictory. While unfortunate, this fact is not surprising.

Before testing any theoretically established relation, the researcher needs to first face a serious preliminary obstacle that cannot be surpassed by conventional financial/statistical methods: that of convincing herself that the *quantitative* measures she intends to use for describing the complex, qualitative CSR concepts, usually supplied by professional data providers (like SAM, KLD, etc.), are adequate and representative, i.e., the numbers in the spreadsheets reflect the degree of implementation of the CSR principles.[1] The answer to the question of the impact of corporate socially responsible behavior on profitability might very well remain elusive due mainly to the difficulties related to measuring CSR performance.

With this important caveat in mind, a first preliminary step in any quantitative study testing any theoretically established relation between corporate socially responsible behavior and profitability consists of evaluating the *association* between firms' CSR scores and their operational performance; i.e., are particularly high/low levels of firm efficiency associated with determined constellations of CSR measures?[2] This step is usually accomplished by a linear regression analysis. Given the multidimensional nature of the CSR measures currently available and due to the possibly weak association between CSR variables and firm performance, special attention is needed when performing this step.

The main goal of our paper is to show how a careful statistical assessment of the association between a broad set of CSR variables and firms' operational performance as measured by return on assets (ROA), assessment based on modern variable selection and non-linear semi-parametric and non-parametric regression techniques, can be carried out.[3] Our analysis

[1]Such studies succeed in investigating the relation between CSR and firm profitability only to the extent to which the scores they use are *relevant* quantitative measures of the operational implementation of the qualitative CSR principles. A CSR dimension might have a significant relation with firm performance even if no impact of the CSR measure used in the study is found. The negative result might only reflect the inadequacy of the proposed measure for the CSR dimension under discussion.

[2]As before, lack of positive/negative findings in this preliminary analysis will not infirm the relevance of a given CSR dimension. It will only establish the absence of association between the particular measure of that dimension as proposed by the data provider and the measure of firm performance.

[3]Since we conceived our study as a "case study," i.e., an in-depth examination of a single instance, we limit our interest to this accounting-based measure of the efficiency of a firm. The methodological approach that we advocate can, of course, be used with other performance measures.

includes a broad range of accounting variables usually used in profitability studies.[4] We extend the commonly used linear framework to allow for non-linearities in the association between the accounting as well as CSR variables and the measure of firm performance. As a by-product of the analysis, we evaluate the relative contribution of accounting variables.

Given the importance of the CSR measures to be used in empirical tests of theoretically-established relations and given the abundance of providers of CSR scores, the outlined methodology can be thought of as a necessary first step in judging and comparing the relevance of different sets of measures of CSR standing. Once performed, the analysis indicates the variables and/or the data sets that might be instrumental in testing theoretically motivated relations.

Our analysis is conducted on a ten-dimensional firm-specific data set covering all three aspects of the CSR principles. The data set was provided by the Sustainable Asset Management (SAM) Group, an independent asset management company specializing in sustainability investments.[5] Firm profitability is measured by ROA.

The contribution of our paper to the literature is two-fold. First, we try to address the multi-faceted nature of the CSR concept by analyzing the properties of a high dimensional set of CSR scores covering a broad range of issues. We believe that a successful investigation of the relation between CSR performance and firm profitability is possible only if based on scores intended to quantify the complex nature of this concept.

Second contribution is methodological in nature. We argue that the methodology of linear regression commonly applied in studies like ours is affected by a number of fallacies (the most serious ones being the choice of the variables to be included in the model and the over-simplification of the relationship to be described through the rigid restriction to linearity). We advocate a more refined approach based on the results from the modern field of statistical learning. In a first step, we implement the classical linear regression but try to improve its performance by using theoretically sound methods of variable selection, such as Breiman's method (Breiman and Spector, 1992) or the early stopping rule in boosting with component-wise linear bases (Bühlmann and Yu, 2003). In a second step, we investigate the performance of non-linear methods such as regression trees and smooth splines.

The findings of the paper are as follows. First, if the CSR variables are included in the pool of explanatory variables, only a small number of scores of the SAM data set (corporate governance, talent attraction, and codes of conduct) seem to show some association with firm profitability as measured by return on assets (ROA). Lower than average corporate governance seems to be associated with lower than average ROA. Higher than average corporate governance, talent attraction, and codes of conduct scores are associated with higher than average ROA.

[4]In the frame of estimating the association between the CSR variables and ROA, these variables should be thought of as control variables.

[5]SAM, in cooperation with the Dow Jones Indexes and STOXX Limited, publishes and licenses the Dow Jones Sustainability World Indexes (DJSI), a series of global sustainability benchmarks launched in September 1999. The indexes are based on SAM's corporate sustainability assessment, which identifies global sustainability leaders on the basis of economic, environmental, and social criteria.

Second, non-linear methods seem to bring significant improvements in explaining profitability. This gain seems to come from accounting for non-linearities in the relation between the classical accounting variables and the performance measure. Including the CSR variables brings more noise in the estimation than gains in explanatory power, i.e., the gaining model is a non-parametric, non-linear regression with only the accounting variables commonly used in profitability studies. Allowing for non-linearities but ignoring the CSR variables seems best.

The paper is organized as follows: Section 2 briefly presents the economic theories on the relationship between CSR and operational performance and Section 3 reviews the relevant literature. Section 4 provides an overview of the profitability measure, the CSR performance measures, as well we the other independent variables suggested in earlier studies. In Section 5, we elaborate on the methodological motivation underlying our choices of statistical tools, described in Section 6. Section 7 describes the data, and our empirical results are contained in Section 8. Section 9 contains a few concluding remarks.

2. Economic Theories about the link between CSR and Firm Profitability

Theoretical ideas on the impact of CSR practices on firm profitability appeared in the management science literature as early as in the beginning of the 20th century (Crane et al., 2008, p. 21). The discussion later developed both in the economic literature with a focus on the cost and revenue implications of CSR practices as well as in the financial theoretical literature where simple stock return implications for investment carried under environmental, social, and corporate governance considerations were put forward.

Since in this paper we are interested in the *operational performance* of firms (the profitability measure we employ is firms' return on assets), we will briefly review the competing economic theories as well as the empirical evidence related to this type of performance. We note that, often, the theoretical arguments concern the environmental dimension.

The first direction of thought includes the views expressed in traditional neoclassical economics based on the increased costs argument. Stringent environmental standards are believed to lead to higher compliance costs for companies in sensitive sectors of the economy (e.g., oil and gas), implying a competitive disadvantage. As the environment is one of the main production factors, imposing limitations on it (e.g., through investments in cutting edge technologies as a way to reduce pollution) will increase costs (Palmer et al., 1995; Siebert et al., 1980).

Alternatively, the revisionist economic view proponents (see Porter and van der Linde, 1995, among others) base their theory on the argument of cost savings and revenue increases. Stringent environmental standards are assumed to generate a competitive advantage when the compliance involves innovative technology. The need to comply with such standards will stimulate new solutions that can yield new technologies that improve resource productivity and production process efficiency and at the same time avoid waste. Also, better technologies implies lower environmental risks as less environmental taxes or charges will be paid and fewer pollution rights will need to be purchased (Schaltegger and Müller, 1998). Moreover, the benefits of environmental awareness might materialize not only through the production process itself but also through increased demand for a firm's products by those who value CSR (Margolis et al., 2007).

Similar arguments can be developed around the social and governance dimensions as well. Regarding the social dimension (or "relation to other stakeholders"), it has been claimed that, while a source of additional costs for the firm, it has potential benefits. For example, investment in human capital might lead to higher operational performance through increased employee motivation and possibly higher worker productivity (Becchetti et al., 2008). Similarly, while consideration toward the community (e.g., charitable giving) is primarily a cost, the reputational benefits could be significant and could materialize in increased sales or in attracting valuable employees.

The impact of governance dimension (or "relation to third parties") on profitability is straightforward as there is no downside to it. Transparent governance and business practices imply lower agency costs as well as better business decisions based on better understanding of the business, which leads to higher profitability.

We want to emphasize that the above-mentioned economic theories make no statements on the linear/non-linear nature of the relationship between measures of CSR standing and firm profitability. Most empirical studies in this area repeatedly estimate and test a linear relationship without contemplating the likely presence of non-linearities. In fact, there is no *economic* reason[6] to suppose a linear relation between the CSR variables and performance. It is in fact very likely that economic performance dependency on CSR is strongly non-linear. It is intuitively pertinent, for example, that investing the same amount in becoming more eco-efficient when one is at the frontier of best practice is likely to yield significantly less in terms of firm profitability than investing the same amount when one is a laggard in the field.

Finally, a word on our choice to look at an operating performance measure of profitability. We believe that the relation between CSR standing and operational profitability is more direct and more stable over time than that with a "pure" market-based performance measure. Market-based measures are by definition influenced by investors' *perceptions* and *expectations*, which, in turn, are affected by their access to information. Thus, confounding factors, like the ability of the financial market to price in CSR information as well as the evolution of CSR disclosure standards, will be partly responsible for the impact of CSR standing on market-based performance measures.[7] Therefore, we believe that, despite well-known shortcomings of accounting measures, an investigation of the relation between CSR standing and operational performance measures will give a more truthful picture of how CSR practices *per se* affect firm performance.

[6]The linear assumption is only motivated by convenience: easy estimation and apparently easy interpretation of the results.

[7]The relation between CSR standing and market-based financial performance can hence be expected to fluctuate possibly only due to information availability issues.

3. Empirical evidence

The empirical results on the relation between corporate responsible practices and profitability do not allow for a clear, strong conclusion. However, Orlitzky et al. (2003) find in a quantitative meta-analysis of 52 studies that social performance and, to a lesser extent, environmental performance are likely to increase corporate financial performance, especially when measured by accounting-based indicators. Margolis et al. (2007) confirm the overall positive association between CSR and financial performance in a meta-analysis of 167 studies, though its magnitude is quite small.

Most empirical studies analyzing the relationship between CSR performance and operating performance use the regression model based on the increasing/decreasing efficiency argument (see Section 2). The analysis is performed at the firm level, with different firm profitability measures being employed, while firm- and industry-specific characteristics are captured through a number of control variables.

A number of methodological limitations are common to most studies. The working assumption is that the firm performance measure and CSR scores are related in a *linear* fashion. Moreover, firm social responsibility performance is almost always uni-dimensional (either as an overall CSR score or, often, as an environmental score). A brief review of the most important studies in the field follows.

Hart and Ahuja (1996) find evidence of a positive effect of environmental performance measures on one and two period ahead return on assets (ROA) and return on equity (ROE) respectively, while no correlation within the same time period is found. Russo and Fouts (1997) also suggest that environmental performance is positively associated with return on assets and that this association is more pronounced for high-growth industries. King and Lenox (2001) find that pollution prevention is associated with higher ROA, while Wagner et al. (2002) document a uniformly negative relation between environmental and financial performance for companies within the pulp and paper industry (through a simultaneous equations framework). There is also some evidence indicating a time-varying premium (first negative then increasingly positive) for "environmental winners," based on cross-sectional data analysis of the relation between Innovest environmental performance measures and both ROA and Tobin's q (Güenster et al., 2005).

van der Laan et al. (2008) is the only study – to our knowledge – to investigate the impact of a variety of CSR measures on two measures of profitability, namely ROA and earnings per share (EPS). CSR performance is measured along a "good" and a "bad" dimension by, respectively, seven strength and seven weakness variables. The study finds that while weakness indicators are generally negatively associated with profitability, i.e., low CSR hurts profitability, positive indicators have no impact, i.e., high CSR performance has no impact. The only exception is the environment strength indicator, which is found to hurt profitability.

4. Corporate Social Performance and our Firm Profitability measure

4.1. **Corporate social performance measures.** Understanding the relationship between CSR standing and firm profitability is made difficult by the lack of consensus on exactly how compliance with CSR should be converted into numbers. Various proxies have been used over the years, ranging from being highly subjective, e.g., surveys of business students (see Heinze, 1976), business faculty members (Moskowitz, 1972), and the Fortune rankings (see McGuire et al., 1988; Herremans et al., 1993; Preston and OBannon, 1997), to being extremely specific, i.e., official corporate disclosures, annual reports to shareholders, and CSR reports as in Hart and Ahuja (1996), Karpoff et al. (1998), and Muoghalu et al. (1990)), respectively, to more standardized and comprehensive measures, i.e., sustainability scores. These scores or rankings represent an aggregate measure of compliance with the principles of CSR and are provided by specialized, independent agencies such as KLD Research & Analytics, Innovest, the SAM Group, etc. The scores are constructed through in-house research based on corporate public and private documentation.

Despite the high level of standardization and the special care applied in the process of constructing the scores, it is almost impossible to determine whether these more sophisticated CSR measures are objective since the original source of information rests with the company itself and few companies have their CSR reports externally verified. Thus, CSR measures are strongly subject to subjective bias and questions about impression management. We approached the choice of a relevant set of scores for our analysis with this caveat in mind.

By choosing the SAM Group, one of the leading institutions specializing in sustainability investments, as our data provider, we tried to capitalize on SAM's many years of sustainability expertise. The relevance of its data set is further demonstrated by the fact that the SAM Group, in cooperation with the Dow Jones Indexes and STOXX Limited, publishes and licenses the Dow Jones Sustainability World Indexes (DJSI), the first global sustainability indexes launched in 1999.

Another important reason for our choice was that the SAM Group designed a comprehensive, ten-dimensional[8] CSR measure. The scores cover all three main aspects of CSR: relation with third parties (the G in ESG), with various stakeholders (the S in ESG), and with the environment (the E in ESG). Eight of the scores are based on firm-provided answers to the SAM questionnaire.[9] The other two, social reporting and environmental reporting, are constructed based only on publicly available information: CSR annual reports, the CSR part of annul reports, company websites, and other references to a company's CSR performance.

[8]The number of dimensions covered in the SAM questionnaire varies over time. Fourteen dimensions were covered in 2002 but only ten in 2006.

[9]The SAM questionnaire is sent out annually to firms that together represent 20% of market capitalization in all industries and in all major economies in the world.

Relations to third parties are quantified by three scores[10]: corporate governance, risk & crisis management, and codes of conduct. Labor practice indicators, human capital development, talent attraction, and corporate citizenship measure corporate relations with various stakeholders in the firm. Finally, relations to the environment are measured by scores on eco-efficiency, defined in terms of efficient use of resources and reduction in greenhouse gas emissions.

Moreover, the SAM Group takes a number of precautions to limit the impact of the mentioned subjectivity criticism, e.g., supportive documentation is provided by firms, predefined multiple-choice questions are used to limit qualitative answers, regular checks are run to verify the truthfulness of the answers, and the process of constructing the scores is highly standardized and externally assessed by Price Waterhouse Coopers. All these measures may to a certain extent alleviate the preoccupation with the subjective bias yet not fully remove it.

Finally, a word on the between-firm comparability along each dimension of the SAM scores (which range from 0 to 100). While the extent to which a firm's engagement with respect to a certain CSR issue can be judged as being higher or lower than that of a another firm might be debatable, especially at a conceptual level, SAM pays particular attention to constructing the scores in such a way to facilitate such a comparison. For example, a firm with a score on corporate governance that is twice as high as that of another firm can be claimed to be devoting (roughly) twice the efforts and resources to enforce corporate governance. While not perfectly monotone measures, SAM scores do allow for a ranking from low to high.

4.2. **Firm profitability measure.** Research studies investigating the impact of compliance with CSR principles on corporate profitability and efficiency usually employ accounting-based measures such as Return on Assets (ROA), return on equity (Becchetti et al., 2008) and return on sales (King and Lenox, 2001). The hypothesis to be tested is that incorporating CSR principles in the business activities might be associated with higher/lower profitability ratios, by the arguments presented in section 2.

We have chosen ROA as profitability measure, defined as net income plus (after tax) interest payments but before preferred dividends per unit of average current and last year's assets. This choice is motivated by the fact that ROA is one of the broadest measures of firm operating performance (Russo and Fouts, 1997; Güenster et al., 2005).

4.3. **Variables that explain firm profitability.** Firm profitability is a central issue in economics. A number of variables have been shown to help explain profitability. Also, some of these variables have been found to have confounding effects on the CSR-profitability relationship as well. In this section, we describe the relevant control variables found in the literature that we include among the independent variables in our analysis.

[10]Each score is obtained by summarizing the answers to a varying number of questions, both qualitative and quantitative in nature. As the questionnaire is quite detailed, we have presented in Appendix D the specific questions as well as more details on the score-building process.

Previous research (Ullmann, 1985; McWilliams and Siegel, 2000) has shown that industry affects firm performance through industry-specific factors such as competitive intensity and economies of scale. Also, the importance of different CSR criteria is not similar across industries, i.e., the environmental dimension is likely to have a much stronger impact on profitability ratios of firms in the heavy industry, relative to firms in the consumer services industry.

Together with industry, firm size is used as an explanatory variable in most studies on profitability. It is hypothesized to be weakly negatively correlated with profitability. Size has been found to be related not only to firm profitability (Ullmann, 1985; McWilliams and Siegel, 2000) but also to a firm's CSR performance. Larger firms seem to implement CSR principles more often, possibly due to the fact that they can afford allocating more resources to the adoption of CSR standards. According to Burke et al. (1986), as they grow, firms attract more attention from stakeholders and hence are under more pressure to comply with CSR principles.

A positive association between firm growth and profitability has been extensively documented (see, e.g., Capon et al., 1990). Usually, the proxy used for firm growth is the average percentage change in sales or assets.

Retained earnings is a measure of the capacity of the firm to reinvest in its core business. In most cases, companies retain their earnings in order to invest them in areas where the company can create growth opportunities; e.g., they may buy new machinery, spend more on research and development (R&D), or pay off debt. A positive relationship between profitability and past retained earnings is hypothesized. A positive correlation is also assumed between retained earnings and corporate social performance, as firms with higher retained earnings have readily available resources for projects leading to improved corporate social performance.

Capital intensity has also been claimed to explain performance (Fama and French, 2000). A possible explanation could be that higher capital intensity should be associated with higher performance because investment in technical capital results in knowledge enhancement, which, in turn, leads to product and process innovation. Ultimately, innovation leads to enhanced efficiency (McWilliams and Siegel, 2000). We operationalize capital intensity by the ratio of capital expenditure to total property, plant and equipment.

Although McWilliams and Siegel (2000) suggest that R&D investment is a non-negligible determinant for both profitability and CSR performance, we follow Waddock and Graves (1997) and control indirectly for R&D level through industry dummies.

Firm risk has also been suggested in previous articles to be a factor that affects both economic and CSR performance (Ullmann, 1985). While the association between firm risk and economic performance has been more widely discussed, the association between firm risk and CSR performance can be explained in various ways. Waddock and Graves (1997) argue that management's risk tolerance influences its attitude toward CSR activities that elicit savings, incur present or future costs or build or destroy markets. The higher the management's risk aversion, the less likely they are to care for CSR activities. Moreover, Dowell et al. (2000) argue that highly risky firms (measured by leverage) are less able to make long-term investments necessary to enhance CSR performance, and therefore a negative association between firm risk and CSR is again expected.

Finally, our sample consists of firms that operate in different fiscal regimes and, more importantly, have different attitudes toward corporate social responsibility (e.g., Japanese firms usually score very low in corporate governance by any measure). In order to account for these differences, we have constructed four broad regional dummies: the Anglo-Saxon countries (the US, Canada, the UK and Australia), Europe, Japan, and the rest of the world. While these constructed regional dummies are not a perfect measure, they are helpful nonetheless.

Note that our analysis, besides its focus on the impact of CSR standing on profitability, will yield conclusions on the nature and strength of the association between these variables and firm profitability as measured by ROA. To our knowledge, our study is one of the few to allow for such a *non-linear* relationship, evidenced by Fama and French (2000). As we will see in Section 8, some of the associations are strongly non-linear, and accounting for their true nature produces significant improvements in explaining ROA.

5. Methodological Motivation

5.1. **Why model selection and model validation are important.** Most quantitative studies on the impact of adopting CSR practices spend considerable efforts assessing the association between measures of corporate social performance and measures of firm profitability. The interest in such a relationship has at least two motivations: to gain explanatory insight into how the economy operates and to improve predictive power when forecasting firms' future profitability.

From a modeling perspective, the success of any method trying to establish such an association depends on its ability to separate between *systematic patterns* present in the data set and *stochastic components*, features that are specific to the sample at hand but that will not show up in an independent sample of the same data. The proposed model should be parsimonious, i.e., contain as few parameters and/or variables as possible, and should not capture any sample-specific details, i.e., should not fit the noise present in the sample.

These goals motivate the observance of two important distinct steps in statistically sound model fitting, involving distinct subsets of the data set. The first step, often called *model estimation*, decides on the significant variables in the model as well as on their contribution and uses a portion of the available data, often called *training sample*, while the second step, the so-called *model validation*, evaluates the fit of the model on a disjoint part of the data set, called the *test sample*, and checks that the model has not over-fitted, i.e., that it has not fitted the noise present in the training sample.

The validation step also offers the possibility to compare how well different models fit the data. Moreover, the two steps of model selection and validation become even more important when facing a large set of weak predictors as is the case with our CSR measures.

Maintaining a clear separation between the data subsets used in the two steps of the model building is essential. Often (and statistically not sound), the two steps are performed on the whole data set, leading to models that fit not only the systematic features of the data but also the stochastic component and accrediting the model with better performance than its true one. The model, optimized to capture also the noise present in the sample at hand, will have less predictive power on an independent data set generated by the same data-generating process (Mosteller and Tukey, 1977; Efron, 1986; Chatfield, 1995).

5.2. **How to correctly estimate, validate and compare models.** One statistically sound approach to model estimation and validation that is also intuitively easy to explain is the so-called *cross-validation* technique. Using this method (see Allen, 1974; Stone, 1974; Geisser, 1975), the data is repeatedly split into two groups of observations, with one group used for model building (the training sub-sample) and the other group set aside for model validation (the test sub-sample).

In more detail, for a given sample, a random partitioning into k roughly equally-sized subsets (k depends on the sample size; typical values are 5 or 10) is done (hence the name k-fold cross-validation). Then, one of the k subsets is left out for validation purposes, i.e., becomes a test or validation sub-sample, while the remaining $k-1$ are used for model building, i.e., they become a training or estimation sub-sample. The model estimated on the training sub-sample together with the independent variables in the test sub-sample are used to predict the dependent variable in the test sub-sample, and an out-of-sample residuals sum of squares (RSS) is calculated. Repeating this procedure k times, k out-of-sample RSS measures are obtained, which are then averaged to produce an overall RSS measure of the model under scrutiny.

Performing the above procedure simultaneously for more than one model constitutes a statistically sound way of judging competing models. This is the approach we will use when comparing various variable selection methods for linear models as well as a number of non-linear models in the data analysis part in Section 8.

6. Statistical Modeling

6.1. **Linear framework.** By far the most commonly employed statistical technique for establishing the association between CSR scores and ROA values is the multiple linear regression analysis based on the assumption of a linear relationship between corporate social performance and profitability. Despite its simplicity, the linear approach, as it is commonly applied in practice, is plagued by a number of problems both in the model building step as well as in the model validation stage.

The most important decision in the model building step of a linear model is that on the variables to be included in the regression. Inclusion of non-significant variables increases the estimation variance of the significant coefficients. Various model selection techniques have been developed (including forward selection, backward elimination, and stepwise and exhaustive search). All these procedures involve repeated tests of significance in their search for the optimal set of explanatory variables, a practice long known (to statisticians) to result in erroneous inclusion of non-significant variables in the final model (see Hocking, 1976; Miller, 1984, 1990).[11] This recognized bias in model selection has stimulated a lot of research in the statistical field, leading to the introduction of improved variable selection techniques. Two such techniques, cross-validation for model selection (Freedman and Peters, 1984; Breiman and Spector, 1992; Shao, 1996) and the early stopping rule of the boosting with component-wise linear bases (Bühlmann and Yu, 2003) will be used in our empirical analysis to address this weakness of the linear approach.

Another well-known problem related to the building of a linear model is its potential instability when the hypothesis of linearity is not realistic. When the relationship in the data is not well described by the linear assumption, removal or addition of a few observations can dramatically change the set of optimal variables as well as the estimated coefficients. This potential weakness will be addressed in the non-linear framework described below.

6.2. **Non-linear framework.** The strong assumption of linearity of the relationship between CSR scores and firm profitability is another important limitation of the existing studies. It is in fact very likely that economic performance depends on CSR in a strongly non-linear fashion (see Section 2). Therefore, our analysis relaxes the assumption of linearity and focuses on two non-linear regression approaches put forth in the last decade in the field of statistical learning. They allow for non-linearities in the relationship between the dependent and the explanatory variables and, as we will see, explain the relationships in the data significantly better.

The non-linear regression methods considered in this paper are, first, smooth splines[12] whose performance is improved by the technique of boosting (Schmid and Hothorn, 2008) and, second, regression trees[13] performance-enhanced by the bagging technique (Breiman, 2001). For details on the techniques of boosting and bagging used to improve the performance of our non-linear models, see Appendix A.

The two combinations of non-linear regression models and performance-enhancing techniques, i.e., boosting the component-wise smooth spline and bagging regression trees, were selected as being among the best performing methods in the statistical learning field that are, at the same time, interpretable. We cannot emphasize enough how differently the two methodologies approach the estimation of the associations in the data.

[11]It is the sequential testing of a large number of related models in search of the optimal one that increases the likelihood of including in the final model predictor variables that have a true random association with the dependent variable.

[12]Since splines are often used in econometric analysis, we will not explain them in detail (for the procedure of boosting component-wise smooth spline, see Bühlmann and Yu, 2003).

[13]Trees explain variation in a single response variable Y by repeatedly splitting the data into more homogeneous groups, i.e., dividing the space of the explanatory variables X into regions on which the response variable Y takes more similar values. As a consequence, some variables can heavily influence the predicted response in some part of the range of the explanatory variables and not be a factor at all in another part of the range. For more details on regression trees, see (Breiman et al., 1984, Ch. 8).

We conclude this section with a few remarks on the issue of variable selection in the case of the non-linear approaches we advocate. Boosting itself functions to a certain extent as a variable selection tool. An early stopping rule, designed to avoid over-fitting, yields also a method of variable selection. This early stopping rule will be used to select relevant variables, both with component-wise linear bases and with smoothing splines.

In the case of bagging, the choice of the relevant explanatory variables is made based on the *relative importance* of the variables as well as on their *partial dependence function*[14] (Friedman, 2001). For more details on these tools, see Appendixes B and C.

To summarize, in both linear and non-linear modeling, we strictly observe the following two modeling principles. First, we make sure that the model evaluation step is not performed on the same data used for the model building (using the same data for performing both operations is the common practice in econometric studies in the field). Second, our assessment of model performance is consistently done via the statistically sound *cross-validation* methodology outlined previously. It is the average Residual Sum of Squares measure produced by cross-validation that yields the ultimate judgement on the performance of a model (be it linear or non-linear), i.e., on the relevance of a given set of explanatory variables. In doing so, we avoid over-fitting, i.e., fitting the noise in the data, and get a realistic picture of the explanatory power of the models compared.

7. Description of the Dataset

As mentioned, the firm profitability measure we consider is Return on Assets (ROA). Among the explanatory variables, we count the SAM scores for ten CSR criteria: social reporting (*SocRep*), environmental reporting (*ERep*), corporate governance (*CGov*), risk and crisis management (*RiskMg*), codes of conduct (*CoCo*), labor practice indicators (*Labor*), human capital development (*HuCap*), talent attraction (*Talent*), corporate citizenship (*CCiti*), and eco-efficiency (*EcoEff*) (a detailed description of these CSR dimensions is included in Appendix D). The scores range from 0 to 100, with 100 being the maximum value.

The process of acquiring and analyzing the company-specific CSR information ends sometime between April and June every year, implying that the CSR scores in year t mostly reflect the CSR standing of the company during year $t-1$. For this reason, ROA corresponds to the fiscal year that ends between June year $t-1$ and May year t, depending on when firms announce their yearly accounting reports. This way, an overlap of at least 6 months in the time period that CSR performance and ROA reflect is assured.

The five other control variables, as discussed in Section 4.3, have been obtained from Datastream as follows: firm size (*Assets*) is measured as log of total assets stated in US dollars (Datastream item 07230), three year average percentage change in sales (*dSales*, hereinafter sales growth rate, item 08633), retained earnings (*Rearn*) is the ratio of retained earnings the

[14]The partial dependence function is defined as the effect of a given explanatory variable on the predicted dependent variable after accounting for the average effects of the other variables.

year before (item 03495) to total assets in the current year (item 02999), capital intensity (*CEFA*) is the ratio of capital expenditures (item 04601) to the value of property, plant, and equipment (item 02501), and firm risk (*Lev*) is measured by the ratio of long-term debt (item 03251) to total assets (item 02999).

Eight industry dummies are also included among the explanatory variables. The nine industries, defined in accordance with the MSCI global industry classification standards (GICS), are: oil & gas, industrials, consumer goods, healthcare, consumer services, telecommunications, utilities, technology, and basic materials. We exclude financials since they have a role of intermediaries in the economy and their balance sheet structure is different from that of other sectors (which may adversely distort the distribution of financial variables). Also, CSR characteristics are structurally different for financials than for other sectors.

Finally, regional and year dummies have been added to the analysis. The geographical distribution of the sample firms is 39% European, 37% US, UK, Canadian and Australian, 20% Japanese, and 4% from rest of the world (African, Latin-American, and the Pacific).

8. Data analysis

Our analysis is performed on SAM's CSR data over five years, from 2002 to 2006. After matching it with economic data, eliminating a few abnormal observations in these variables (most likely due to errors in data recording in Datastream or unusual firm behavior), and eliminating financials, the sample size is reduced to 1882 firm-year observations[15]. The histograms of ROA, the five control variables, and the ten CSR scores are shown in Figure 1.

We note that some scores seem more informative than others. For example, the scores for codes of conduct, corporate governance, labor practices, and talent follow an almost normal distribution. Few scores seem less informative; e.g., most of the companies got perfect (or almost perfect) scores for environmental reporting and social reporting. As such, these criteria carry very little or no information and we expect to not see them among the variables that help explain profitability. Other scores like corporate citizenship and (to a lesser extent) eco-efficiency seem to be uniformly distributed between their minimum and maximum values.

The distribution of some of the non-CSR explanatory variables (retained earnings or capital intensity) shows positive skewness.

In the regression analysis that follows, all explanatory variables are centered and standardized such that they have sample variance equal to one. This operation facilitates the interpretability of the results.

[15]Outlier evaluation was done as follows: ROA (in %) larger than 40 (2 observations lost); ROA smaller than -25 (7 observations lost); log(Assets) smaller than 12 (2 observations lost); long-term debt to assets ratio larger than 1 (1 observation lost); 3 years average change in sales (in %) larger than 70% (12 observations lost); capital intensity ratio larger than 1 (25 observations lost); negative ratios of retained earnings to assets were put to zero, instead of elimination, due to their high relative incidence – 7.8% of the data; ratio of retained earnings to assets larger than 1 (11 observations lost).

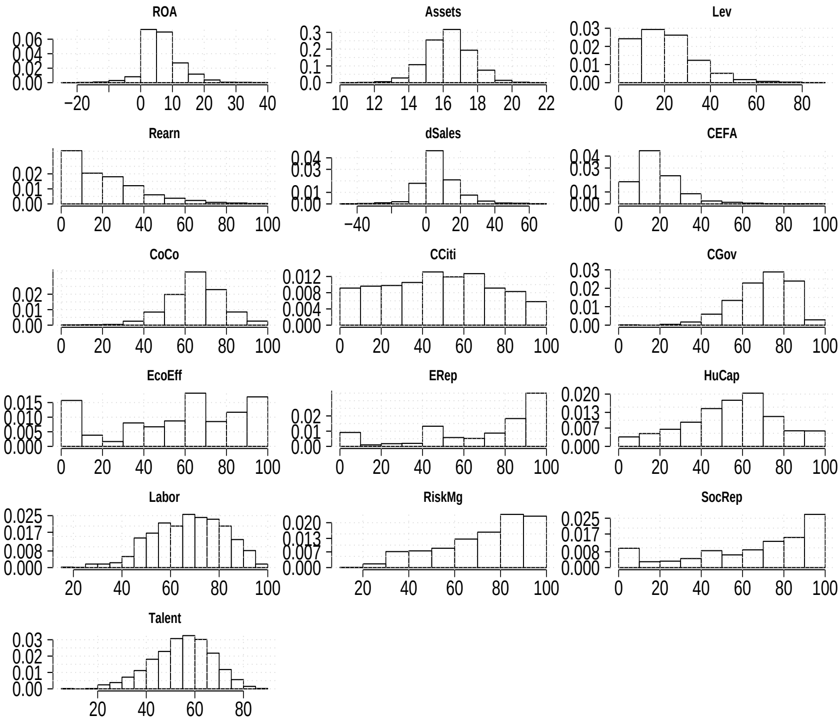


Figure 1. Histograms of the dependent variable ROA (first graph), of the control variables except dummies (next 5 graphs), and of the CSR scores for non-financial firms 2002-2006.

Several methods for choosing the relevant explanatory variables, both linear and non-linear, have been used. Their relative performance has been evaluated through a *cross-validation exercise*, described in subsection 5.2.

8.1. **Linear framework.** We began the data analysis assuming a linear relation between firm profitability as measured by ROA and the explanatory variables discussed in Section 7. We then performed a linear regression analysis and made use of three different methods of variable selection: the common stepwise algorithm, Breiman's variable selection approach, and the selection provided by the early stopping rule of boosting.

The performance of these three methods was evaluated based on the reduction in average residuals sum of squares (RSS) over the five validation samples relative to two benchmarks, i.e., two linear regression without any variable selection and wi th different sets of explanatory variables. The first regression included all the variables at hand, both CSR and control variables (described in Section 8), while the second one included only the control variables.[16]

[16]The cross-validation exercise is explained in detail in Section 5.2.

According to the RSS measures, variable selection does not improve the performance of any of these two linear models.[17]

Thus, under the assumption of a linear relationship between profitability and both CSR and non-CSR explanatory variables, adding SAM's CSR variables to the set of explanatory variables does not bring any gains in explaining ROA even when sound statistical methods of variable selection are used. A linear regression including only the accounting and dummy variables explains the data equally well as one where SAM's CSR variables are included.

As discussed earlier, the assumption of linearity has no real economic base and is commonly made for statistical convenience. Next we will investigate the relationship between profitability and the explanatory variables while relaxing the assumption of linearity.

8.2. **Non-linear framework.** Two non-linear regression approaches were considered for modeling the relation of interest: smooth splines and regression trees. Their performance was improved by applying bagging in the case of the tree modeling and boosting to the component-wise smooth spline.

8.2.1. *Contribution of non-linearity and CSR variables.* To better judge the relevance of the CSR variables, the non-linear analysis will focus on two sets of explanatory variables. We will evaluate models that allow for non-linearities in the relationship between ROA and, first, the accounting and dummy control variables only and, second, the whole set of explanatory variables including SAM's CSR variables. As usual, the performance of the models is judged in terms of average RSS values produced by cross-validation.

Non-linear regressions capture the association between ROA and the explanatory variables significantly better than the linear approach. Bagging regression trees provides a reduction of around 15% in average out-of-sample RSS for both sets of explanatory variables, while boosting with smooth splines implies a more modest (roughly) 4% reduction (Table 1, Column I, Row I and II).

Explanatory variables	(I) Both CSR and non-CSR		(II) Only non-CSR	
Benchmark Model, Explanatory Variables	Boosting Spline	Bagging Trees	Boosting Spline	Bagging Trees
(I) linear regression, both CSR and non-CSR	-4.3%	-14.8%	-	-
(II) linear, only non-CSR variables	-3.9%	-14.4%	-5.0%	-16.9%
(III) non-linear model, only non-CSR	1.1%	3.0%	-	-

TABLE 1. Percentage change in out-of-sample residuals sum of squares measure relative to various benchmarks (Row I to III), for boosting with component-wise smooth splines and bagging with regression trees, when both CSR and non-CSR variables are included in the model (Column I), and when only non-CSR variables are included (Column II). A negative number means improved performance.

A closer look at Table 1, Column I, Row III reveals an important finding: non-linear regressions with only the accounting control (non-CSR) variables yield the best performance.

[17]The RSS slightly increases by between 0.4% with stepwise regression and almost 2% with Brieman's method.

The performance of both non-linear approaches slightly worsens when the CSR variables are added to the set of explanatory variables. Two implications follow from this. First, it shows that even traditional accounting variables known to explain profitability are non-linearly associated with profitability. Second, it indicates that SAM's CSR variables do not contribute to explaining profitability.

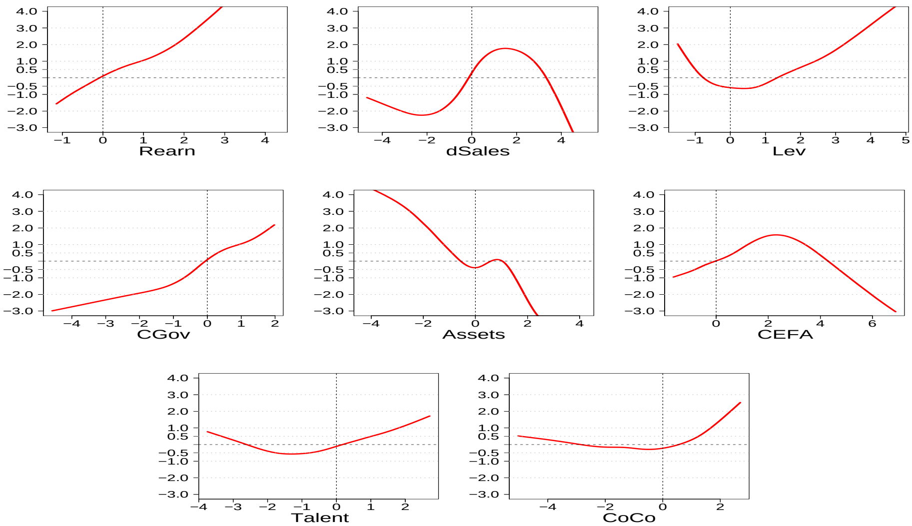
8.2.2. *CSR effect on ROA.* Given that the performance (as measured by the average RSS) of the two sets of explanatory variables, i.e., the accounting and dummy control variables and the whole set of explanatory variables including SAM's CSR scores, is comparable, it makes sense to investigate the shapes of the relationships between ROA and the CSR explanatory variables that are found relevant by the estimation. The first two rows of graphs in Figure 2 display the partial dependence functions for the eighth relevant variables selected by boosting with component-wise smooth splines. Since the explanatory variables are centered and standardized, the abscissa negative (positive) values correspond to a below-average (an above-average) performance. The dependent variable is also centered so that negative (positive) values correspond to a penalty (reward) relative to average profitability.

The lower half of Figure 2 displays the partial dependence functions for the relevant explanatory variables selected by bagging regression trees. The variables are chosen according to the ranking in the graph of relative importance in Figure 3 as well as based on the magnitude of their partial dependence functions. The contribution of the variables ranked lower than eighth place is negligible. Please note that the eight most influential variables in explaining firm profitability identified by the method of bagging trees have also been chosen by the boosting smooth spline approach (shown in Figure 2).

The graphs in Figure 2 show that the shapes of the partial dependence functions estimated by the two extremely different non-linear methods are very similar, with the exception of the shapes at the tails of the distributions (above and below two standard deviations from the mean). Moreover, the magnitude of estimated impact is also similar across the two methods with the exception, again, of the distribution tails. At the ends of the range of the explanatory data, boosting with smoothing splines produces explosive upwards or downwards curves for partial dependence functions, which is in sharp contrast to those estimated by regression trees. One reasonable explanation resides in the rigid, smooth nature of the splines that adapt with difficulty to changes in the shape of the relation in the distribution tails where fewer data points are available. Hence the tendency of estimated shapes to extrapolate patterns that are present more into the center of the distribution, causing a growth off-scale in the ends of the range of the explanatory data.

As the method of bagging regression trees seems to both improve out-of-sample prediction the most and be more stable with respect to extreme observations, our interpretation of the results puts most weight on the graphs in the last two rows of Figure 2.

Only three – out of ten – CSR variables, i.e., corporate governance, talent attraction, and codes of conduct, have been selected as relevant in the non-linear framework. Generally,



Partial dependence of ROA on the explanatory variables estimated with component-wise smoothing spline. The method yields a reduction of 4.3% in average out-of-sample residuals sum of squares with respect to a linear regression with all variables.

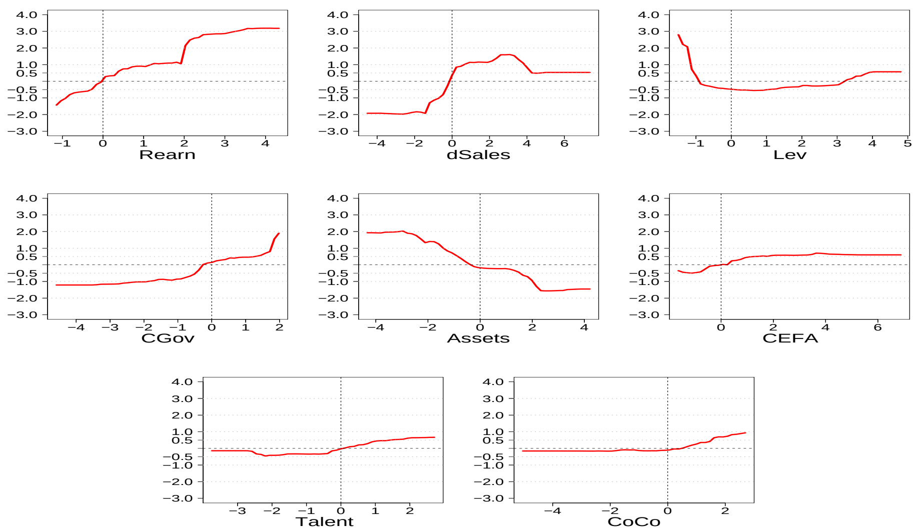


FIGURE 2. The partial dependence functions of ROA on the explanatory variables estimated by bagging trees. The method yields a reduction of 14.8% in average out-of-sample residuals sum of squares with respect to a linear regression that uses all explanatory variables. See Section 7 for definitions of the names of the CSR scores.

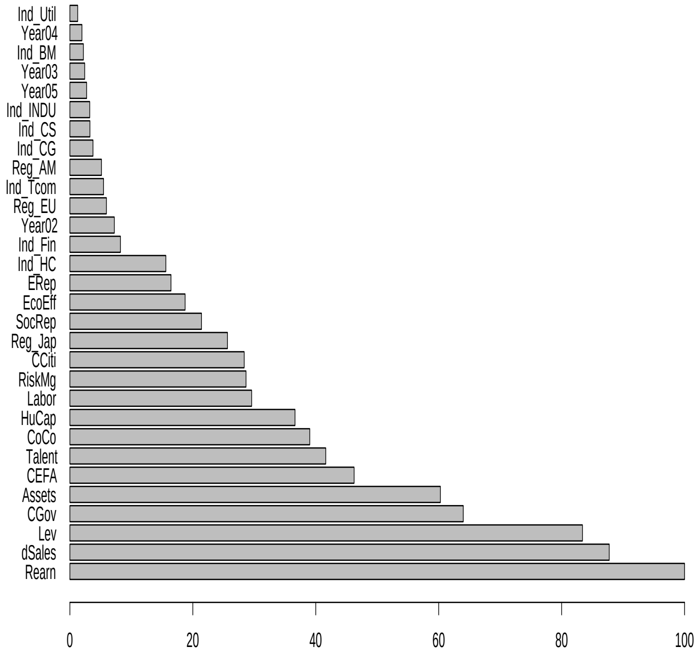


FIGURE 3. Relative Variable Importance Plot for bagging regression trees. See Section 7 for definitions of the names of the CSR scores.

below-average CSR scores are associated with lower ROA and above-average scores are associated with higher ROA, yet the relationship is far from linear. Within this group, corporate governance has the strongest impact on ROA, ranging from -1% to 2%, while the impact of talent attraction and codes of conduct is much smaller,i.e., from -0.5% to 1%.

It was of course expected that the five non-CSR control variables (retained earnings, sales growth rate, leverage, size, and capital intensity) would show the highest estimated impact on profitability. However, corporate governance turns out to place fourth, ahead of firm size and capital intensity (Figure 3). For corporate governance scores ranging from one standard deviation below to one standard deviation above the mean, ROA varies concavely between -1% and +0.50%. However, the shapes of the well-above- and well-below-average scores are not symmetric. Exceptional corporate governance is associated with significantly higher profitability (1 – 2% higher), yet there is no distinction between modestly poor and very poor governance in terms of ROA penalty (of -1%).

Unlike governance, talent attraction and codes of conduct have modest impacts on ROA, more visible only on the positive half of their standardized distributions. In this region, the relationship is increasing with ROA, either in a concave manner for talent (up to 0.5% higher) or monotonically (up to 1% higher) for codes of conduct. For scores below the mean, the reaction in ROA is constant almost everywhere, slightly lower for codes of conduct, and 0.5% lower for talent. Thus, above-average engagement with talent attraction or codes of conduct is associated with higher profitability, but engagement below average is only weakly penalized with lower profitability.

Recall that these partial dependence functions were estimated based on the five years of pooled data. This served us well in our attempt to reveal a *stable* relationship between CSR and profitability. As a robustness check, we have also estimated the relationship between ROA and the entire set of explanatory variables on a yearly basis. These yearly shapes, although more wiggly and with more extreme jumps at the range end of some explanatory variables, are qualitatively identical to the pooled ones.

8.2.3. *Impact of non-CSR variables on ROA.* The estimated impact of the non-CSR variables on profitability often confirms intuition. As suggested by the RSS evaluation, all shapes are strongly non-linear, yet sales growth rate and leverage show the strongest signs of non-linearity.

The variables with the largest impact on ROA are retained earnings (*Rearn*) and sales growth rate (*dSales*). For moderate above- and below-average values in these variables, ROA increases almost monotonically from -2% to 2%. As expected, having been profitable in the recent past and having retained part of that generated income for investment in productive projects renders the firm profitable in the next period as well. Also, sales growth rate is a proxy for a firm's growth opportunities and therefore predicts profitability. Non-linearity is more visible for well-above-average values. Extremely high *Rearn* is associated with substantially higher profitability gains of 2 – 3%. Moderately above-average values in *dSales* produce most profitability gains of more than 1.5%. As for more extreme values, the benefits diminish to 0.5% higher ROA.

Leverage (*Lev*), the third most important variable in terms of strength of association with ROA, has a negative effect on ROA for below-average values. Very small leverage can improve ROA by as much as 3%, but the effect levels off at lower 0.5% ROA for both below- and above-average values. However, the effect turns positive again for very high leverage levels. As anticipated, firm size (*Assets*) and capital intensity (*CEFA*) are, respectively, negatively and weakly positively related to ROA for above- and below-average values. The effects level off at +/-2% for well-above-average values in size and at 0.5% for extreme positive values in *CEFA*.

Finally, a note on industry, regional, and year dummies. The variable importance plot (Figure 3) seems to indicate that these (otherwise important) dummies contribute very little to explaining ROA. However, it is the flexibility of a non-linear relationship that allows to implicitly model differences in ROA due to such classifications through equivalent differences embedded in its estimated non-linear relationship with the other (quantitative) explanatory variables.

9. Conclusions

We believe we have demonstrated that non-linear regression methods bring important improvements in explaining profitability. Our findings are strong evidence that an increased level of statistical sophistication is needed if one wants to uncover relationships explaining firm profitability. The relevance of our findings is buttressed by the fact that two very different non-linear approaches produce similar results as they deliver identical sets of relevant explanatory variables as well as significant reductions in the average RSS, a statistically sound measure of model performance. We believe that increased attention from the academic community to non-linear approaches might well be a source of important achievements.

We find that SAM's CSR scores do not seem to help explain firm profitability as measured by ROA. However, including them in the pool of explanatory variables worsens the performance of the model only marginally.

The variables with the greatest explanatory ability are those capturing either firms' growth opportunities or their competitive advantages. Such variables have first-order effects on profitability. By contrast, CSR scores can only have smaller-order effects on profitability and therefore do not produce major improvements in prediction error.

If one decides[18] to include SAM's CSR scores in the set of explanatory variables, we find that only a small number of CSR scores like corporate governance, talent attraction, and codes of conduct seem to have some association with ROA. In particular, corporate governance places ahead of known explanatory variables like firm size and capital intensity.

The positive effects of corporate governance on profitability confirm something we already knew from earlier studies (Gompers et al., 2003). The effects of talent attraction and codes of conduct, on the other hand, have been harder to evaluate. Our study shows that above-average scores in these dimensions are associated with up to 1% higher than average profitability, while below-average scores slightly hurt ROA. This finding resonates with Edmans (2008), who showed that a portfolio of "100 Best Companies to Work for in America" had a significantly long-term higher financial performance, in terms of stock returns, than its benchmark.

Most of SAM's CSR scores considered in the analysis do not seem to contribute to explaining profitability. The "big absent" is the environment dimension. Contrary to many empirical results indicating some (negative or positive) association between the CSR environmental dimension and firm performance (King and Lenox, 2001; Konar and Cohen, 2001), our analysis shows no relation whatsoever.

These negative results could be due to many reasons, e.g., sample composition. A larger and more diverse sample would have allowed for a more sensitive analysis and might have shed light on contributions that, with the data at hand, are impossible to detect. The construction of the scores could provide, of course, another possible explanation. Aggregated differently,

[18]The decision can be motivated by the fact that the performance is only slightly worse than that of a model with only the accounting variables.

the answers to the questionnaires might have produced more informative scores. Or it could be that other statistical techniques can individuate better the set of relevant variables.

We finish by restating the important caveat emphasized in the Introduction. Our findings do not suggest that the qualities these scores set out to measure are not important. It is only to the extent that the SAM scores correctly measure the operational implementation of CSR concepts and that the return on assets truthfully reflects the profitability of a firm that the conclusions of our analysis carry to the the understanding of the "true" association between CSR and profitability. We have most likely not yet found the best way to measure CSR, and the results of the analysis might constitute a stimulus (for researchers as well as for sustainability-rating companies) to rethink the methodology behind the construction of such scores.

Appendix A. Bagging and boosting

Bagging and boosting are both methods that can be applied to large classes of regression models (linear or non-linear) with the aim of improving their accuracy. Both originate from the machine learning research community. Both methods operate by estimating the regression model of choice – which can be unstable, as the linear model is – many times on different subsamples of the original data set (resampling). The improvement in performance comes from averaging in the case of bagging and through systematically targeting parts of the space where the modeling seems less accurate in the case of boosting.

In bagging, a number of training sets are constructed by re-sampling the original data. The idea of bagging (also called "bootstrap aggregation") is simple: many bootstrap samples are drawn from the available data, some prediction method is applied to each bootstrap sample, and then the results are averaged to obtain the overall prediction (Breiman, 1996a). Averaging reduces the variance. Bagging works best when the base regression procedure that is being bagged is not very stable. That is, when small changes in the sample lead to significant differences in the predictions obtained using a specified method, bagging can result in a significant reduction in average residuals sum of squares. (See Breiman, 1996b, ,for additional information about instability.)

Boosting (see Freund and Schapire, 1997) assigns a set of different weights to the original data when re-sampling to produce training sets (the higher the weight, the more likely it is that the data point will be part of the new training set). The weights are adjusted after each estimation of the model on a training set, i.e., by increasing the weights of the observations that are not well-explained by the last-estimated model and by decreasing the weight of those that are well explained by the estimated relationship. The method sequentially fits the data that has not been well-fitted so far. The final model is an aggregation of the consecutive models proposed.

Bagging (and to a lesser extend boosting) can be viewed as a way of exploiting the instability of the regression model to improve its explanatory accuracy. In various comparisons with bagging, boosting often but not always does better, leading researchers in the field of machine learning to favor boosting over bagging. Moreover, there is strong empirical evidence that boosting is generally resistant to over-fitting.

A.1. **A More Formal View on Boosting.** Boosting (Hastie et al., 2001, p. 303-305) is a way of fitting an additive expansion in a set of elementary "basis" functions. Generally, it takes the form

$$f(x) = \sum_{m=1} \beta_m b(x; \gamma_m) \tag{1}$$

where β_m, $m = 1, \ldots, M$ are the expansion coefficients and $b(x; \gamma)$ are usually simple functions of the multivariate argument x, characterized by a set of parameters γ. The structural properties of the boosting function estimator are induced by a linear combination of the structural characteristics of the basis functions.

To make the estimator (1) precise, one needs to specify the basis functions and a modality of estimation of the coefficients β.

Forward stage-wise modeling approximates the solution of (1) by sequentially adding new basis functions to the expansion without adjusting the parameters or the coefficients of those that have already been added. More concretely, the steps are

(1) Initialize $f_0(x) = 0$.
(2) For $m = 1$ to M
   (a) Compute
$$(\beta_m, \gamma_m) = \operatorname*{argmin}_{\beta,\gamma} \sum_{i=1} L(y_i, f_{m-1}(x_i) + \beta b(x_i; \gamma)).$$
   (b) Set
$$f_m(x) := f_{m-1}(x_i) + \beta_m b(x_i; \gamma_m).$$

The loss function L will be, for us, the squared-error loss

$$L(y, f(x)) = (y - f(x))^2. \tag{2}$$

Under this assumption, the previous algorithm turns into

(1) Initialize $f_0(x) = 0$.
(2) For $m = 1$ to M
   (a) Compute the residuals
$$U_i = Y_i - f_{m-1}(X_i), \quad i = 1, \ldots, n.$$
   Fit the residual vector $U_1, \ldots, U_n$ to $X_1, \ldots, X_n$ (by some type of regression) to estimate (β_m, γ_m).
   (b) Set
$$f_m(x) := f_{m-1}(x_i) + \beta_m b(x_i; \gamma_m).$$

In other words, the L_2-boosting amounts to refitting residuals multiple times.

Estimation of f with boosting is usually done by pursuing an iterative steepest descent of the empirical risk $n^{-1}\sum_{i=1^n L(Y_i, f(X_i))}$. To fully specify our approach, we need to still define the basis functions to be used in the empirical analysis. They correspond to component-wise linear least squares and component-wise smoothing spline. More formally, they are

$$\hat{\beta}^{(\mathcal{S})} x^{(\mathcal{S})} \tag{3}$$

$$\hat{\beta}^{(j)} = \sum_{i=1}^{n} X_i^{(j)} U_i / \sum_{i=1}^{n} (X_i^{(j)})^2, \quad \mathcal{S} = \operatorname{argmin} \sum_{i=1}^{n} (U_i - \hat{\beta}^{(j)} X_i^{(j)})^2 \tag{4}$$

and

$$h^{(\mathcal{S})}(x^{(\mathcal{S})}) \tag{5}$$

$$h^{(j)} = \underset{h}{\operatorname{argmin}} \sum_{i=1}^{n} (U_i - h(X_i^{(j)}))^2 + \lambda \int (h^{''})^2 dx, \quad \mathcal{S} = \operatorname{argmin} \sum_{i=1}^{n} (U_i - h^{(j)}(X_i^{(j)}))^2,$$

where λ is a tuning parameter.

In every iteration, the basis (3) selects the best one predictor variable in a simple linear model (not necessarily a different one in each iteration) in the sense of ordinary least square fitting. The method is also known as matching pursuit. The component-wise smoothing spline (5) belongs to the class of generalized linear models. The best predictor variable is chosen and its partial contribution to f is estimated by a least squares smoothing spline. The degrees of freedom in the smoothing spline base procedure are chosen small such as df=4. This choice yields low variance but possibly large bias, which can typically be removed by additional boosting iterations.

Appendix B. Relative importance of input variables

The first tool introduced (Friedman, 2001) is the relative influences I_j, of the individual inputs X_j, on the variation of $f(\mathbf{x})$ over the joint dependent variable distribution.

$$I_j = \left(E\left[\frac{d\hat{f}(\mathbf{x})}{dx_j}\right]^2 \operatorname{var} X_j \right)^{1/2} \tag{6}$$

For a given sample and a regression tree T estimate of f, a sample version $\hat{I}_j(T)$ of the quantity (6) can be introduced (see Friedman, 2001). For a collection of regression trees $\{T_m\}_{m=1}$, as for example, the result of boosting or bagging, i.e., for a family of regression trees that in a linear manner yield an estimate of f, the sample version of I_j is constructed taking a linear combination of the ones corresponding to the individual trees:

$$\hat{I}_j = \frac{1}{M} \sum_{m=1} \hat{I}_j(T_m) \tag{7}$$

Appendix C. Partial dependence function

Visual representation is one of the most powerful interpretational tools. Unfortunately, it is limited to low-dimensional arguments. Viewing functions of higher-dimensional arguments is more difficult. It is therefore useful to be able to view the partial dependence of the approximation $f(\mathbf{x})$ on selected small subsets of the input variables. Although a collection of such plots can seldom provide a comprehensive depiction of the approximation, it can often produce helpful clues, especially when $f(\mathbf{x})$ is dominated by low-order interactions.

Consider the sub-vector $\mathbf{X}_\mathcal{S}$ of $l < p$ of the input predictor variables $\mathbf{X} = (X_1, X_2, \ldots, X_p)$, indexed by $\mathcal{S} \in \{1, 2, \ldots, p\}$. Let $\mathcal{C}$ be the complement of $\mathcal{S}$. A general function $f(\mathbf{X})$ will generally depend on all of the input variables: $f(\mathbf{X}) = f(\mathbf{X}_\mathcal{S}, \mathbf{X}_\mathcal{C})$. The *partial dependence function* is defined as the effect of $\mathbf{X}_\mathcal{S}$ on f after accounting for the average effects of the other variables $f(\mathbf{X}_\mathcal{C})$:

$$f_\mathcal{S}(\mathbf{X}_\mathcal{S}) = E_{\mathbf{X}_\mathcal{C}} f(\mathbf{X}_\mathcal{S}, \mathbf{X}_\mathcal{C}). \tag{8}$$

A sample version of (8) can be defined as

$$f_\mathcal{S}(\mathbf{X}_\mathcal{S}) = \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{X}_\mathcal{S}, \mathbf{X}_{i\mathcal{C}}), \tag{9}$$

where $\mathbf{X}_{1\mathcal{C}}, \ldots, \mathbf{X}_{N\mathcal{C}}$ are the values of $\mathbf{X}_\mathcal{C}$ occurring in the data sample. For additive tree models (like the one produced by boosting or bagging), the partial dependence functions are averaged over the constituent trees.

Appendix D. SAM CSR database

The process of obtaining company-specific scores consists of several steps. First, the companies asked to participate in the annual review (these companies represent at least 20% of the market capitalization in all industries and in all major economies of the world) fill out a general questionnaire covering the areas described below. The companies need to support the information disclosed with appropriate documentation.

Next, the answers to all questions are converted into grades. Two averaging processes follow. In the first one, the scores for each of the ten criteria listed below are obtained (these are the scores that we analyze). A final company-specific score describing the firm's overall CSR standing is also produced by averaging the ten grades obtained in the previous step. This final score is then used to identify the leaders in sustainability which are included in the Dow Jones Sustainability Indexes.

The ten CSR scores are based on the information outlined below (taken mainly from SAM Questionnaire in 2009, which used to be available at $http://www.sustainability-indexes.com/07\_htmle/assessment/criteria.html$).

Social Reporting and **Environmental Reporting** reflect content, context, and coverage of the social/environmental reporting included in firm reports or websites (e.g., own publication, part of a sustainability/CSR report or annual report) compiled by a SAM industry analyst.

Relation to third parties:

(1) **Corporate Governance** concerns information on the number of members of the Board of Directors (BD); information on the number of employee/trade union representatives, required by law, that are on the BD/Supervisory Board (SB); information on how many Board members have executive functions; whether the Board is headed by a non-executive and independent chairman and/or an independent lead director; information on the functions for which the Board explicitly assumes formal responsibility, e.g., strategy, audit accounting, risk management, selection/nomination/remunaration of board members and top management, CSR/corporate citizenship, and sustainable development; whether the BD/SB has issued a formal corporate governance policy; information on the percentage of the main nationality represented on the company's BD relative to all other nationalities represented; how many women are members of the company's BD/SB; information on the percentage of non-audit related (e.g., for management consulting) fees paid to the company's auditing firm; and whether the company communicates the remuneration/compensation of its BD/SB's members and other highest paid senior directors/executives externally.

(2) **Risk & Crisis Management** concerns information on the chief risk officer/person responsible at the group level (name, position, and reporting line); information on the person responsible for issue/reputation management (i.e., coordination and communication of issues with high potential risk to the company's reputation) at the group level (name, position, reporting line and number of levels from the BD/Executive Board); information on which elements are included in the company's crisis/emergency plans (i.e., business continuity plan, communication with the media and other critical audiences/stakeholders affected, co-ordination between departments involved (e.g., Public Relations, Investor Relations, Manufacturing, Customer Service, Finance and Risk Management departments), frequent rehearsal/testing of plans, mechanisms for early internal/external notification of an emergency situation).

(3) **Codes of Conduct** concerns for which areas corporate codes of conduct have been defined at the group level (including subsidiaries), i.e., corruption and bribery, discrimination, confidentiality of information, money laundering and/or insider trading/dealing, security of staff/business partners/customers, environment/health/ safety; what mechanisms are in place to assure effective implementation of the company's codes of conduct (responsibilities, accountabilities and reporting lines are systemically defined in all divisions and group companies; dedicated help desks; codes of conduct linked to employee remuneration; employee performance appraisal systems integrates compliance/codes of conduct; disciplinary actions in case of breach, e.g., zero tolerance policy;

compliance system is certified/audited/verified by third party); which aspects are covered by the company's anti-corruption and bribery policy at the group level (including subsidiaries), i.e., bribes in any form, direct or indirect political contributions, political contributions publicly disclosed, charitable contributions and sponsorship, charitable contributions and sponsorship publicly disclosed; information on the percentage of coverage of the company's codes of conduct and anti-corruption and bribery policy relative to the total number of employees group or worldwide/contractors, suppliers, service providers/subsidiaries/joint ventures; whether the company publicly reports breaches (e.g., number of breaches, cases etc) of its codes of conduct/ethics and anti-corruption and bribery policy.

Relation to Stakeholders:

(1) **Labor Practice Indicators** concern information on which performance/management indicators the company uses regarding: non-discrimination/diversity (i.e., % females of total workforce, % female managers in total workforce, breakdown of workforce based on minority, culture, or similar), equal remuneration female/male (i.e., average salary male and female at executive level, management level, and non-management level), freedom of association (i.e., % of employees represented by an independent trade union or covered by collective bargaining agreements, consultations or negotiations with trade unions over organizational changes), layoffs (e.g., number of employees laid off in the last fiscal year, consultations or negotiations with employees over organizational changes – such as restructuring and outsourcing), and health and safety (e.g., tracking of safety performance, tracking of work-related fatalities, tracking of near misses or similar crisis events); which systems are in place to collect and handle employee grievances and complaints to ensure that workers can raise their concerns in confidentiality (e.g., help line, whistleblowing policy, independent person or department in charge of solving complaints by employees such as diversity committee or company ombudsman, counseling, strict confidentiality ensured, policies and related information widely circulated in appropriate languages); whether the company publicly endorses (having signed or publicly acknowledging adherence to) one or more of the following charters/frameworks: UN Universal Declaration of Human Rights, ILO Tripartite Declaration of Principles concerning Multinational Enterprises and Social Policy, OECD Guidelines for Multinational Enterprises, other national/international charters related to labor practices/basic rights issues).

(2) **Human Capital Development** concerns whether the company measures and controls the long-term success of its human resource policies in a formal/standardized way using for example qualitative operating indicators/ratios, non-financial indicators/ratios, cost-based indicators/ratios, investment- or value-based financial indicators, and human resource-based indicators; whether the company has a medium-term workforce and skills plan comparing current employees and their skills with the future

number, type, and skills of employees required to execute the business plan (i.e., for business/performance units generating more than 75%/50-75%/less than 50% of revenue); the percentage of skilled employees and executives receiving a regular formal evaluation of their performance; information on how senior/middle management is appraised (i.e., regular performance appraisal by line superior, multidimensional performance appraisal, systematic use of agreed measurable targets and indicators, and formal comparative ranking of managers); the percentage of employees who follow a company training programs specific to their job category (e.g., sales manager) before or at the onset of their job/position.

(3) **Talent Attraction** concerns the percentage of employees hired based on a validated recruitment process/selection test in the last fiscal year; the percentage of skilled employees (managerial, professional, and technical employees) having left the company in the past year relative to the total average number of skilled employees during the same period; percentage of the workforce that are systematically re-assigned within the company or in extreme cases outplaced due to weak individual performance relative to the average total workforce during the last fiscal year; whether the company regularly tracks and benchmarks employee satisfaction against industry peers with regard to rewards and recognition, leadership, supportive/collaborative team environment, personal development possibilities, job satisfaction/opportunity to make a difference, working environment (health and safety, social climate, etc.), identification with corporate values and strategy, other; information on the satisfaction level of the company's employees this year relative to last year based on the company's employee satisfaction surveys (i.e., higher, constant, decreased, no trend); information on the share of performance-related compensation for each employee category as a percentage of total compensation (excluding pension plans and fringe benefits) that the company paid out in the last year; information on the percentage of variable compensation that is based on corporate and/or individual performance for top/senior management, middle/lower management, sales staff, and technical specialists, as well as overall company average; information on the type and its percentage share of total performance-related compensation (excluding pension plans and fringe benefits) that the company pays out/grants, i.e., stocks or other forms of stock-related compensation (e.g., options), other long-term compensation (not directly stock-related), profit shares (or similar), sales or order commission (or similar), bonus pool based on profit, divided based on management assessment, scorecard target bonus set in relation to salary granted on the basis of management assessment, other; information on group-wide employee benefits provided by the company in addition to government schemes (i.e., pension plans, health and/or accident insurance for employees, medical care for employee families, disability insurance/programs, maternity and/or paternity leave, child care, flexible work schemes, employee assistance program, other); whether the company offers the

choice of supplementary private pension plans with a sustainability/socially responsible component to its employees (and which).

(4) **Corporate Citizenship** concerns information on whether the company has a system in place to systematically measure the impact of its contributions in order to further improve/re-align the company's philanthropic/social investment strategy (i.e., business outcomes and impact, social outcomes and impact, impact on corporate reputation and stakeholder satisfaction, other); an estimate of the monetary value (in % of pre-tax profit/EBITDA) of the company's philanthropic contributions/voluntary social investments (excluding marketing and advertising budgets) in cash, in-kind giving (employee volunteering during paid working hours/product or service donations/projects or partnerships or similar).

Relation to Environment: **Eco-Efficiency** summarizes information on the company's reduction targets, trend, and performance with respect to total direct GHG emissions, total water use, total energy consumption, and total waste generation.

References

Allen, D. (1974). The relationship between variable selection and data augmentation and a method for prediction. *Technometrics*, 16(1):125–127.

Becchetti, L., Giacomo, S. D., and Pinnacchio, D. (2008). Corporate social responsibility and corporate performance: evidence from a panel of us listed companies. *Applied Economics*, 40(5):541–567.

Breiman, L. (1996a). Bagging predictors. *Machine Learning*, 24(2):123–140.

Breiman, L. (1996b). Heuristics of instability and stabilization in model selection. *The Annals of Statistics*, 24(6):2350–2383.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. (1984). *Classification and Regression Trees*. Chapman and Hall.

Breiman, L. and Spector, P. (1992). Submodel selection and evaluation in regression; the x-random case. *International Statistical Review*, 60(3):291–319.

Bühlmann, P. and Yu, B. (2003). Boosting with the l2 loss: Regression and classification. *Journal of the American Statistical Association*, 98(462):324–339.

Burke, L., Logsdon, J., Mitchell, W., Reiner, M., and Vogel, D. (1986). Corporate community involvement in the san francisco bay area. *California Management Review*, XXVIII:122–141.

Capon, N., Farley, J., and Hoenig, S. (1990). Determinants of financial performance: A meta-analysis. *Management Science*, 36(10):1143–1159.

Chatfield, C. (1995). Model uncertainty, data mining and statistical inference. *Journal of the Royal Statistical Society*, 158(3):419–466.

Crane, A., McWilliams, A., and Matten, D. (2008). *The Oxford Handbook of Corporate Social Responsibility*. Oxford University Press.

Dowell, G., Hart, S., and Yeung, B. (2000). Do corporate global environmental standards create or destroy market value? *Management Science*, 46(8):1059–1074.

Edmans, A. (2008). Does the stock market fully value intangibles? employee satisfaction and equity prices. *Working paper Series at the R.L. White Center for Financial Research*, 16-07(http://papers.ssrn.com/sol3/papers.cfm?abstract_id=985735):[Accessed 19.04.2010].

Efron, B. (1986). How biased is the apparent error rate of a prediction rule? *Journal of the American Statistical Association*, 81(394):461–470.

Fama, E. and French, K. (2000). Forecasting profitability and earnings. *The Journal of Business*, 73(2):161–175.

Freedman, D. and Peters, S. (1984). Bootstrapping a regression equation: Some empirical results. *Journal of the American Statistical Association*, 79(385):97–106.

Freund, Y. and Schapire, R. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55:119–139.

Friedman, J. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, 29(5):1189–1232.

Geisser, S. (1975). The predictive sample reuse method with applications. *Journal of the American Statistical Association*, 70(350):320–328.

Gompers, P., Ishii, J., and Metrick, A. (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics*, pages 107–155.

Güenster, N., Bauer, R., Derwall, J., and Koedijk, K. (2005). The economic value of corporate eco-efficiency. *European Financial Management, no*, $[doi : 10.1111/j.1468-036X.2009.00532.x]$.

Hart, S. and Ahuja, G. (1996). Does it pay to be green? an empirical examination of the relationship between emission reduction and firm performance. *Business Strategy and the Environment*, 5(1):30–37.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The elements of statistical learning: data mining, inference and prediction.* Springer, New York.

Heinze, D. (1976). Financial correlates of a social involvement measure. *Akron Business and Economic Review*, 7(1):48–51.

Herremans, I., Akathaporn, P., and McInnes, M. (1993). An investigation of corporate social responsibility reputation and economic performance. *Accounting, Organizations and Society*, 18(7-8):587–604.

Hocking, R. (1976). The analysis and selection of variables in linear regression. *Biometrics*, 32(1):1–49.

Karpoff, J., Lott, J., and Rankine, G. (1998). Environmental violations, legal penalties, and reputation costs. *Working paper*, Available at: $http://www.law.uchicago.edu/files/files/71.Lott\_.Environment.pdf$:[Accessed 10.08.2010].

King, A. and Lenox, M. (2001). Does it really pay to be green? an empirical study of firm environmental and financial performance. *Journal of Industrial Ecology*, 5(1):105–116.

Konar, S. and Cohen, M. (2001). Does the market value environmental performance? *The Review of Economics and Statistics*, 83(2):281–289.

Margolis, J., Elfenbein, H., and Walsh, J. (2007). Does it pay to be good? a meta-analysis and redirection of research on the relationship between corporate social and financial performance. *Working paper*, Available at :$http : //stakeholder.bu.edu/2007/Docs/Walsh, JimDoesItPaytoBeGood.pdf$:[Accessed 10.08.2010].

McGuire, J., Sundgren, A., and Schneeweis, T. (1988). Corporate social responsibility and firm financial performance. *The Academy of Management Journal*, 31(4):854–872.

McWilliams, A. and Siegel, D. (2000). Corporate social responsibility and financial performance: Correlation or misspecification? *Strategic Management Journal*, 21(5):603–609.

Miller, A. (1984). Selection of subsets of regression variables. *Journal of the Royal Statistical Society*, 147(3):389–425.

Miller, A. (1990). *Subset Selection in Regression*. Chapman and Hall, New York.

Moskowitz, M. (1972). Choosing socially responsible stocks. *Business and Society Review*, 1(1):71–75.

Mosteller, C. and Tukey, J. (1977). *Data Analysis and Regression*. Addison-Wesley, Reading, Massachusetts.

Muoghalu, M., Robinson, H., and Glascock, J. (1990). Hazardous waste lawsuits, stockholder returns, and deterrence. *Southern Economic Journal*, 57(2):357–370.

Orlitzky, M., Schmidt, F., and Rynes, S. (2003). Corporate social and financial performance: a meta-analysis. *Organization Studies*, 24(3):403–441.

Palmer, K., Oates, W., and Portney, P. (1995). Tightening environmental standards: the benefit-cost or no cost paradigm. *The Journal of Economic Perspectives*, 9(4):119–132.

Porter, M. and van der Linde, C. (1995). Green and competitive: ending the stalemate. *Harvard Business Review*, 73(5):120–134.

Preston, L. and OBannon, D. (1997). The corporate social-financial performance relationship: A typology and analysis. *Business and Society*, 36(4):419–429.

Russo, M. and Fouts, P. (1997). A resource-based perspective on corporate environmental performance and profitability. *Academy of Management Journal*, 40(3):534–559.

Schaltegger, S. and Müller, K. (1998). Calculating the true profitability of pollution prevention. In Benett, M. and James, P., editors, *The Green Bottom Line: Environmental Accounting for Management; Current Practice and Future Trends*. Greenleaf Publishing.

Schmid, M. and Hothorn, T. (2008). Boosting additive models using component-wise p-splines. *Computational Statistics and Data Analysis*, 53(2):298–311.

Shao, J. (1996). Bootstrap model selection. *Journal of the American Statistical Association*, 91(434):655–665.

Siebert, H., Eichberger, J., Gronych, R., and Pethig, R. (1980). *Trade and environment: a theoretical enquiry*. Elsevier Scientific Publishing Company, Amsterdam.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society*, 36(2):111–147.

Ullmann, A. (1985). Data in search of a theory: A critical examination of the relationships among social performance, social disclosure, and economic performance of u. s. firms. *The Academy of Management Review*, 10(3):540–557.

UNEPFI (2007). Responsible investment in focus: How leading public pension funds are meeting the challenge.

van der Laan, G., van Ees, H., and van Witteloostuijn, A. (2008). Corporate social and financial performance: an extended stakeholder theory, and empirical test with accounting measures. *Journal of Business Ethics*, 79(3):299–310.

Waddock, S. and Graves, S. (1997). The corporate social performance- financial performance link. *Strategic Management Journal*, 18(4):303–319.

Wagner, M., Phu, N. V., Azomahou, T., and Wehrmeyer, W. (2002). The relationship between the environmental and economic performance of firms: an empirical analysis of the european paper industry. *Corporate Social Responsibility and Environmental Management*, 9(3):133–146.

Paper II

Strategic Corporate Social Responsibility and Economic Performance

Constantin Belu[a] and Cristiana Manescu[bc]

Abstract

This paper proposes a novel Corporate Social Responsibility (CSR) index based on a Data Envelopment Analysis (DEA) model. Acknowledging the argument that companies might favor those CSR dimensions that provide strategic competitive advantages, we argue that the index can capture companies' strategic approach to CSR. Furthermore, our findings reveal a neutral relationship between this strategic CSR index and economic performance as measured by ROA and Tobin's Q, when controlling for firm unobserved heterogeneity and past economic performance. By contrast, an equally-weighted index of the same CSR indicators is found to be negatively related with ROA, which reinforces our claim that this specific DEA-based index is a measure of strategic CSR.

Keywords: Corporate Social Responsibility, Data Envelopment Analysis, Strategic CSR, Difference-GMM

JEL: C23, C67, M14

[a] Dept. of Economics, University of Gothenburg, Box 640, 405 30, Sweden. Tel: +46 317 862 516; E-mail: constantin.belu@economics.gu.se.

[b] Centre for Finance- Department of Economics, University of Gothenburg, Box 640, 405 30, Sweden. Tel: +46 317 864 412; E-mail: cristiana.manescu@economics.gu.se.

[c] Cristiana Manescu gratefully acknowledges financial support received through the research programme Behavioral Impediments to Sustainable Investment, founded by the Swedish Foundation for Strategic Environmental Research, MISTRA.

1. Introduction

There has been a continuing interest in what has become known as Corporate Social Responsibility (CSR) for at least 50 years. Despite disagreements over an appropriate definition, CSR is generally viewed as corporations' responsibility to integrate environmental, social, and governance (ESG) practices into their business model, *beyond mandatory legal requirements*. Moreover, CSR is often associated with the notion of sustainable development.

The increased interest in ESG issues has stimulated a rapid development of empirical literature (Crane et al., 2008) focusing on the relationship between corporate social performance and financial performance (measured by either accounting or market-based variables). While the results are generally mixed, marginally more studies seem to identify a positive, though generally weak, association between the two (Margolis et al., 2007).

One major difficulty in conducting such empirical research, i.e., when analyzing the link between CSR and economic performance, as well as one of the reasons for obtaining conflicting results, lies in defining adequate and representative *quantitative* measures for the complex CSR concept.[1] This paper will address this problem through the construction of a more comprehensive aggregate measure of CSR. Waddock and Graves (1997) expressed the need for a multidimensional measure of CSR applied across a wide range of industries and larger samples of companies and actually designed a set of weights for the various dimensions, based on the views of a panel of experts. We argue that our constructed CSR index meets these requirements and at the same time accounts for the strategic decisions made by managers who bear in mind the ultimate goal of profit maximization.

Strategic CSR is a concept whose origins can be traced back to Baron (2001), who coined the term to refer to a profit-maximizing corporate strategy that can be regarded as socially responsible by some. Burke and Logsdon (1996) also adopted a view similar to strategic CSR, but focused on the corporate strategy attributes that could be linked to CSR. More recently, Siegel and Vitaliano (2007) performed an empirical investigation concerning the determinants of strategic CSR and also reported evidence of economic benefits derived from strategic CSR. While not in a strategic CSR framework, Elsayed and Patton (2005) presented dynamic panel

[1] There are a number of additional reasons why it is challenging to find empirical support for the link between economic performance and CSR. For a more detailed exposure, see Waddock and Graves (1997), UNEP Finance Initiative-Mercer Report (2007), and Belu (2009).

data estimates for the link between environmental performance and companies' financial performance, arguing that very few studies have controlled for firm heterogeneity or considered dynamic effects in the financial/environmental performance relationship. The present paper addresses this problem by employing a Difference GMM estimation framework, as in Blundell and Bond (1998).

Porter and Kramer (2006) make a strong case for strategic CSR, arguing that companies should favor a strategic approach to CSR, i.e., they should identify the corporate agenda that can bring the greatest competitive benefit. They claim that "…the more closely tied a social issue is to a company's business, the greater the opportunity to leverage the firm's resources, and benefit society." Moreover, they argue that companies should carefully select the social issues that intersect with their particular business, because "No business can solve all of society's problems or bear the costs of doing so (...). Other social agendas are best left to those companies (…) that are better positioned to address them."

Recognizing the pertinence of their arguments, we proceed to construct CSR indices that account for the differences between the business models of companies, even within the same industry. In order to achieve this, we resort to Data Envelopment Analysis (DEA), a versatile non-parametric management tool widely used for assessing the relative performance and efficiency of individual decision-making units (DMU) (e.g., firms, schools, hospitals). The DEA feature that we are exploiting most in our study is the assignment of firm-specific sets of weights. These weights are the outcome of an optimization process that seeks to award the most favorable set of weights to each company, with higher weights for outputs where one particular company tends to perform better and lower weights for outputs where the company underperforms, relative to the performance of the other firms in the sample. In contrast to this approach, the current practice when constructing CSR indices is to award a subjectively chosen *a priori* set of weights to various CSR dimensions, the same for all companies in a sample or portfolio.

Please note that the strength of the CSR index we propose does not depend on the number of or the specific underlying CSR dimensions on which it is based. It resides rather in its ability to identify firms that achieve a lot with respect to CSR *relative to their peers* and *given a certain set of CSR dimensions*.

The contributions in this paper are the following: First, we develop an endogenous CSR index that accounts for *strategic* corporate social behavior. Second, we explore the impact of our newly defined measure of CSR on firm performance, measured by return on assets and Tobin's Q which are modeled as autoregressive processes. We therefore control for past economic performance that might influence both current values of economic performance as well as current CSR. We also control for firm-specific effects that have been shown to affect the relationship between CSR and economic performance (Elsayed and Patton, 2005; Baron et al., 2009). Moreover, our empirical model to some extent deals with concerns of causality in the relationship between CSR and firm performance, since it includes both unobserved firm-specific effects and past economic performance.

Descriptive statistics of our constructed strategic CSR index indicate that, as expected, in consumer durable and non-durable experience goods sectors more firms embrace strategic CSR than in any other consumer oriented sector. This confirms the hypothesis that CSR can be used to reduce information asymmetry between consumers and producers, especially for goods whose properties cannot be assessed before purchase (Siegel and Vitaliano, 2007). However, contrary to this hypothesis, we also find that the sector showing largest proportion of firms that behave CSR strategically is basic resources, which could be due to the higher environmental pressure that firms in this sector face, as a result of their environmental impact.

Using the outlined dynamic panel modeling framework, we find evidence that strategic CSR has a neutral impact on firm performance, when controlling for a number of other relevant factors. By contrast, an equally-weighted measure of CSR, based on the same underlying CSR scores, is found to have a persistently negative impact on return on assets. These contrasting findings emphasize the role played in empirical research by the specific aggregation technique of CSR indicators, as well as indicate that our DEA-based method can capture the strategic component of CSR. That the relationship was not found to be positive, as suggested in theory, might be due to the fact that strategic CSR is undertaken only to the extent that marginal cost equals marginal benefit.

The remainder of the paper is organized as follows: Section 2 discusses the heterogeneous nature of the CSR concept and how DEA can be used to construct strategic CSR indices. Section 3 describes our empirical strategy while Section 4 describes the data set. Section

5 presents and discusses our empirical findings as well as several robustness checks. Section 6 concludes the paper.

2. The CSR paradigm. Constructing an aggregate CSR measure

2.1 Difficulties with current CSR measures

There is an ongoing discussion about the appropriate definition of CSR. However, most of the proposed definitions[2] agree that CSR is a multidimensional concept, which is an aspect of particular concern in this paper. Multi-dimensionality implies that various distinct aspects of the nature of a business need to be considered simultaneously when assessing a firm's CSR performance. These distinct criteria are very often clustered into three main subgroups: environmental, social, and governance related.

The methodologies developed by various CSR-rating agencies or data providers[3] involve a subjective weighting of the CSR dimensions' importance. For instance, KLD Research & Analytics, a leading CSR-rating agency, bases its rating criteria on seven qualitative areas: community, diversity, employee relations, human rights (ascribed to the social dimension), environment, governance, and product safety. Furthermore, they detail the above criteria into *strengths* and *concerns*. Their ratings do not involve numbers, but rather qualitative descriptions noted with pluses and minuses.[4]

Sustainable Asset Management (SAM)[5] uses a broader list of criteria for assessing CSR performance and updates it every year. For example, in 2007 the following criteria were rated: corporate governance, risk & crisis management, codes of conduct/corruption & bribery, environmental performance, environmental reporting, labor practice indicators, human capital development, talent attraction & retention, corporate citizenship/philanthropy, and social reporting.

[2] One definition proposed by the European Commission (2001) is as follows: "CSR is a concept whereby companies integrate social and environmental concerns in their business operations and in their interaction with stakeholders on a voluntary basis."
[3] Innovest, IRRC (Investor Responsibility Research Center), Asset4, Sarasin&Cie, KLD Research & Analytics, and Sustainable Asset Management are a few examples.
[4] For details, see the KLD methodology at *http://www.kld.com/research/ratings_indicators.html*
[5] Sustainable Asset Management is a Swiss-based asset management company that computes and updates the Dow Jones Sustainability Index. In addition to general CSR criteria, SAM also computes sector-specific criteria. See *http://www.sam-group.com/* for details.

SAM computes a score from 0 to 100 for each dimension, where 100 means maximum performance. The assessment of the score is done by in-house specialists based either on questionnaires completed by companies or on publicly available sources of information. In the present study, we will use the individual scores provided by SAM to construct our proposed CSR index.

One can notice from the criteria listed above that a wide range of issues are addressed simultaneously in an assessment of a firm's social responsibility. However, there might be differences in the way these issues affect different businesses. Some dimensions are certainly important for some businesses, while others are less relevant. For instance, oil and mining companies are very exposed to environmental risks and therefore deploy strategies with respect to environmental performance accordingly; banks and financial institution put a higher emphasis on risk and crisis management, while IT companies have extensive human capital development strategies and consequently are expected to score high in terms of labor practices. We assume that most managers carefully select the CSR issues that are deemed relevant for their company and then concentrate their efforts in those particular areas.

Different CSR dimensions imply different costs and might provide different benefits and opportunities for profit depending on the nature of the firm's core business. Thus, it is difficult to construct an aggregate measure of CSR in a fair manner, even if accurate information about the achievements in terms of each particular dimension is available. One has to decide on a set of weights to be used for computing an aggregate index. Depending on the structure of the weighting system, more emphasis might be placed on some dimensions and less on others. This subjective way of computing CSR indices is prone to criticism, as it might favor some dimensions over others and therefore some companies over others.

Another research strategy is to conduct separate analyses for each CSR component, i.e., to analyze the association between a measure of economic performance and one particular CSR component (e.g., corporate governance, environmental performance, or labor practices). While many authors have chosen this route (Hart and Ahuja, 1996; King and Lenox, 2001), the potential findings are not really relevant in the CSR context since even if a negative relationship is found between an individual CSR component and economic performance, this should not imply that that particular component/measure should be overlooked when pursuing a CSR agenda. On the other hand, if positive relationships are found, it is difficult to argue that the result can be generalized for companies with significantly different businesses, and it is also unclear how much of its available resources a company should commit to that particular CSR dimension, perhaps to the detriment of other relevant CSR dimensions.

In addition to the above arguments, Baron (2001) argues that in the presence of opportunities for strategic CSR, a positive correlation between economic performance and CSR should be expected. However, when altruism rather than profit maximization drives CSR, a negative relationship might also be possible. Consequently, the empirical analyst should know beforehand whether displayed CSR is a result of altruism, profit maximization, or a threat by an activist. Assuming a generally accepted framework to measure CSR, it would be difficult to distinguish between these types of CSR. For example, Baron (2009) disentangles the strategic component of CSR based entirely on judgment, by selecting those strengths in the KLD database "…that correspond to activities that appear to favor the public directly and seem to be cast that way by the media." The remaining strength indicators are considered to measure a different type of CSR, i.e., responses to social pressure CSR. Our approach avoids such burdensome information requirements and subjective bias since this information is implicitly embedded in our constructed CSR indices based on DEA.

The present paper makes an empirical contribution to the recent strand of literature that distinguishes between various types of CSR: strategic, in response to a threat, and altruistic CSR. Much of this literature is normative and qualitative, mainly consisting of specific predictions based on increasingly complex theoretical models (Besley and Ghatak, 2006; Bagnoli and Watts, 2003; Baron, 2009). Empirical evidence is still scarce.

2.2 How can DEA be used to construct endogenous CSR indices?

The following brief presentation of the DEA method is based mainly on Cooper et al. (2000). The standard DEA models account for production-like processes, where multiple inputs are combined and transformed into several outputs. The main purpose of DEA is to construct an index (score) of relative (to the other units) performance. To obtain this, the first step is to construct a virtual input and a virtual output for each DMU by using a set of (unknown *ex ante*) weights:

$$\textit{Virtual input} = v_1 x_{1o} + \ldots + v_m x_{mo}$$

$$\textit{Virtual output} = u_1 y_{1o} + \ldots + u_s y_{so} \,,$$

where v and u are weights and x and y are inputs and outputs, respectively.

Please note that in our study we only have one input, which is unity, following the arguments presented in Lovell and Pastor (1997). DEA models are *input-oriented* or *output-oriented*. The distinction comes from the way adjustments are made to inefficient units in order to obtain their efficient projections. If adjustments are made in the input space, we have the *input-oriented* approach. If adjustments are made in the output space, we have *output-oriented* models. There is also a third choice, namely models that simultaneously adjust both outputs and inputs, the so-called *additive* models with their *slack-based* variants. However, in our case, the distinction is less relevant. For convenience, we will choose the output-oriented model, like in Banker et al. (1984). Given that there is only one input, the nature of returns to scale is not important either. Lovell and Pastor (1999) showed that an output-oriented model (which assumes constant returns to scale, like in Charnes et al., 1978) with a single constant input, coincides with the Banker et al. (1984) model (which allows for variable returns to scale) with a single input.

The next step is to determine the weights, using mathematical programming techniques, so as to maximize the ratio (θ) between the aggregated virtual output and the virtual input. Consequently, the optimal weights may vary from one DMU to another. Hence, deriving the optimal weights from data is an objective process, compared to fixing them in advance. It is important to realize that there is a ratio for each DMU, so we will get N maximal values (θ^*) and an optimal set of weights for each unit in the sample. The ratio θ^* is restricted to be less than or equal to 1 (or larger than or equal to one in the output-oriented models), and the units that have a $\theta^* = 1$ after optimization are considered to be efficient, i.e., performing best. The lower the calculated θ^* (or the greater than one, in output-oriented models), the more inefficient the unit. This ratio will constitute the base for the CSR index that will be used as an explanatory variable in the empirical model.

DEA constructs the weights endogenously by allowing them to be determined as part of an optimal solution to a formal aggregation problem. More precisely, DEA assigns higher weights to dimensions where a company performs well and lower weights to dimensions where it performs less well. The weights will be chosen such that each company will be placed in the most favorable position in relation to all other companies in the sample. In this manner, we can obtain a score for the relative performance in terms of CSR, for each particular company. The optimal set of weights is determined as part of an optimization process and is company specific. In other words, the DEA weighting system favors dimensions where the company performs better, corresponding to the business strategy implemented by its manager. This means that CSR dimensions that provide competitive advantages and implicitly receive increased ex ante efforts from the manager, as reflected in correspondingly higher SAM marks, will weigh heavier in the aggregate CSR index.

Moreover, DEA provides the means to identify in which dimension a particular firm is lagging behind best practice in CSR terms. It can also provide precise quantitative qualifications to the sub-optimal level for a firm; hence it gives the percentage by which a particular sub-optimal firm should improve in a certain dimension in order to achieve best practice.

Finally a word on the originality of using DEA to construct a strategic CSR index: being a tool extensively used to measure efficiency, DEA is one of the favorite methodologies for measuring environmental performance embedded in firm economic efficiency (see, e.g., Färe et al., 1989). However, to the best of our knowledge, no study has used DEA in the wider context of CSR where it fits particularly well given the necessity to somehow assign weights to the CSR dimensions.

2.3 DEA-constructed strategic CSR index

For our empirical analysis, we construct a DEA index that considers all CSR dimensions as outputs. No particular quantity is considered as an input. We will base our approach on the model developed by Lovell and Pastor (1997), where only one constant input is considered. The reason for this is that we consider each firm as a stand-alone unit, without explicitly accounting for various inputs involved in obtaining the current environmental, social, or governance-related accomplishments. While it is obvious that achieving a satisfactory CSR level might require material inputs, it is usually not clear how these are converted into CSR scores. What we aim to measure is the commitment of a particular firm to the CSR requirements.

If we let $y_j = (y^{1j}, y^{2j} \dots y^{8j})$ represent the vector of CSR scores (provided by Sustainable Asset Management) for the firm j, $j=1 \dots N$ where N is the number of firms in the sample, then we can write the following optimization problem:

$$\max_{\theta,\lambda} \theta$$
$$\textit{subject to}$$
$$\theta y^{ij} \leq \sum_k \lambda_k y^{ik}, i = 1,..,8$$
$$\lambda_k \geq 0, k = 1,...j..., N \quad ,$$
$$\sum_k \lambda_k = 1$$

where i indexes the CSR dimension, k indexes the firms under scrutiny, and λ_k are the assigned weights for each dimension. This model was proposed and used in Lovell and Pastor (1997) to analyze the operating performance of branch offices of a large financial institution in the context of target setting. In our case, we do not have any target requirements, although this procedure can be implemented by the screening agent in a fairly easy manner, as shown in the paper just mentioned.

Moreover, the separation into CSR-efficient and CSR-inefficient firms is performed at the industry level in order to reduce inter-industry heterogeneity, although a CSR index could be meaningfully computed industry-wide. By comparing the CSR standing of firms that are very similar in terms of business models, customer targets, and product classes, one can obtain an as close to strategic CSR measure as possible. Therefore, our DEA-based CSR index is best able to capture firm-specific CSR strategic behavior when looking at very homogeneous industrial sectors.

The DEA-based CSR index returns various degrees of inefficiency, ranging from very close to very far from the frontier. Please note that in an output-oriented model, the computed DEA scores are bound to be larger than one. Moreover, the degree of inefficiency does not vary linearly with the DEA score, and hence a meaningful transformation is required. In order to exploit in the best possible way the limited variation in the data, we construct, based on the industry and year-specific DEA scores, a CSR index with three values: a value of two for efficient firms (DEA score equal to 1), a value of one for slightly inefficient firms (DEA score between 1 and 1.1), and a value of zero for substantially inefficient firms (DEA score larger than

1.1 – and lower than 2, i.e., the maximum by definition). Thus, *the higher our constructed CSR variable, the more CSR-efficient* the firm. The cutoff point of 1.1 was a subjective and debatable decision but was made to ensure enough observations in each group.

One condition has to be met in order for DEA to work reasonably well. The number of firms has to be sufficiently larger than the number of outputs (i.e., the number of CSR dimensions) on which the frontier is built, or else there is a risk that all firms will appear on the frontier. Moreover, please note that the DEA scores are relative to the sample on which they were built, e.g., to a particular industry, each year. Hence, cross-industry inefficiency comparisons of DEA scores cannot be performed, i.e., it cannot be concluded that firms in two different industries with identical inefficient DEA scores are equally inefficient.

In order to provide a minimum evaluation of our constructed CSR index, we will also use an alternative measure of corporate social behavior based on the same underlying CSR scores, yet aggregated differently. It consists of an index that places equal weights on all CSR dimensions, i.e., their equally-weighted mean.[6] Our DEA-based index provides a transparent way of judging one firm's CSR achievements against another's in a similar line of business by placing more weight on the dimensions in which each firm does (relatively) better. This measurement would capture the *strategic* nature of CSR, by emphasizing those dimensions that attract more efforts from a company's management. On the other hand, a simple average of the various CSR scores, which implies that each dimension is equally relevant, should be less likely to measure the strategic component of CSR and more strongly reflect *any non-strategic components* of CSR, i.e., CSR motivated by altruism or by threats by an activist. We will then proceed to investigate the empirical properties of these two CSR measures by analyzing their relationship to firm performance.

Firm performance is expected to be positively related to strategic CSR, since the motivation of embracing strategic CSR is to increase profits, and negatively related to other forms of CSR, since they would increase the costs of the firm (Baron, 2001). The effects of strategic CSR on a firm's economic performance can materialize through several channels: consumer reward, employee and supplier reward, and investor reward Baron et al. (2009), all due

[6] Of course, one can think of many ways of evaluating our constructed CSR index. Its direct competitors would be the indexes built with weights subjectively chosen by analysts at the ranking agencies (which most often are confidential). However, before proceeding to the more complex indexes, we would like to confirm that there are benefits that our DEA CSR index brings relative to a simple average score, not least as it also incorporates a strategic dimension to CSR.

to the fact that various stakeholders value the CSR that the firm provides. *Consumers* might be willing to pay a premium for the goods and services of a firm that provides CSR, while *employees* might be more productive or even accept lower wages. Also, a firm providing CSR might attract better-skilled employees, or better services and products through the *supply chain* from firms that also value CSR. Finally, *investors* may be willing to pay a premium for shares in a firm engaged with CSR. Since the data does not allow separate identification of the individual effects of rewards by consumers, investors, and employees and other suppliers of factor inputs, a combined rewards effect is estimated through the marginal effect of CSR on *EcPerformance* in equation (1).

3. Empirical strategy

Our empirical exercise is mainly concerned with investigating the impact that different measures of CSR have on firm performance. As a proxy for economic performance, we will use both return on assets (ROA_{it}), a profitability measure that expresses the amount of net income plus (after tax) interest payments but before preferred dividends per unit of average current and last year's assets, as well as Tobin’s Q, a forward-looking (expected profits) performance measure that is less prone to managerial manipulation. Tobin’s Q is defined as the ratio of firm market value to the replacement cost of its assets, and we follow Baron et al. (2009) and measure it as the sum of market value of common stock, book value of preferred stocks, book value of long-term debt, and book value of common liabilities, divided by book value of total assets.

Our measures of socially responsible behavior are primarily the DEA-based CSR index described in Section 2.3 as well as an equally-weighted index. By design, the DEA-based CSR index is expected to capture more of the strategic aspect of CSR, while the latter is a better measure of altruistic CSR.

Along the lines of Fama and French (2000), who model ROA as a mean-reverting process, we use a dynamic model for both ROA and Tobin’s Q where their current values are also linked to their first-lag values. Earlier studies (e.g., Elsayed and Patton, 2005; Baron et al., 2009) have also modeled the data-generating process for Tobin’s Q as an autoregressive of order 1 process. Accounting for past values of economic performance also brings the advantage of partially correcting for the potential endogenous nature of CSR performance. If financially successful firms are more likely to undertake CSR activities, by controlling for past performance we implicitly correct for this effect as well (Elsayed and Patton, 2005). Formally, the model is:

$$\begin{aligned} EcPerformance_{it} &= \alpha + \beta_1 EcPerformance_{it-1} + \beta_2 CSR_{it} + \beta_3 X_{it} \\ &+ \sum_{j=1}^{4} \gamma_j DY_j + \sum_{k=1}^{8} \delta_k DInd_k + \sum_{l=1}^{3} \theta_l D\mathrm{Re}\, g_l + (u_i + \varepsilon_{it}) \end{aligned} \quad (1)$$

,

where $EcPerformance$ is either ROA or Tobin's Q, CSR is either the DEA-based CSR index or a simple average CSR index, X is a list of control variables specific to each economic performance measure, ε_{it} is the disturbance, distributed as $N(0,\sigma^2)$, and the Greek letters are parameters.

Following previous studies (e.g., Manescu and Starica, 2007), we consider a number of additional control variables to explain the cross-sectional variation in ROA, such as: firm size, measured as the natural logarithm of assets expressed in US dollars; firm risk, expressed as long-term debt/total assets; capital intensity, calculated as the ratio of capital expenditures to property, plant, and equipment; and sales growth, expressed as a 3-year change in sales (King and Lenox, 2001). In addition to these variables, the set of controls for Tobin's Q also includes dividend to book ratio (Baron et al. 2009) and firm age (Guenster et al., 2005).

Furthermore, economic performance is subject to three types of variation that may be independent of the operations and CSR activities of firms. The first includes factors such as macroeconomic conditions, general market sentiment, and political risks that can affect overall profitability ratios and market values. The second consists of industry-specific factors such as increased or decreased profitability due to shifts in demand or restrictions on supply (Baron, 2009). The third includes factors related to regional variation in accounting reporting standards or consumer preferences for CSR, which may differ, e.g., between Continental European and Anglo-Saxon countries. These types of variation are taken into account by, respectively, year fixed effects (DY_j), industry fixed effects ($DInd_k$), and regional fixed effects ($D\mathrm{Re}\, g_l$) (through dummy variables).

Moreover, as our sample covers firms that differ in terms of, e.g., productivity and management competence in the form of unobserved firm heterogeneity that is constant through time, we need to include time-invariant firm-specific effects, i.e., u_i, in the empirical model.

When estimating a dynamic panel data model, the lagged dependent variable (as an explanatory variable) is *positively* correlated with the fixed-effects term entering the compound disturbance (u_i), which makes the OLS estimator inconsistent with an upward-bias. The Within Groups estimator eliminates this source of inconsistency by removing from each individual the mean values within its group, which eliminates the fixed effect. But for panels with a short time dimension, this transformation introduces a non-negligible *negative* correlation between the transformed lagged dependent and the transformed error term, which will lead to a downward-biased inconsistent Within Groups (i.e., Fixed Effects) estimator. The fact that these two estimators are likely to be biased in opposite directions is useful for evaluating a third consistent candidate estimator, which should lie between these two (Bond, 2002). The general approach relies on instrumental variable (IV) estimators in the General Method of Moments framework. It consists of first differencing the data to eliminate the fixed effects and then using second or earlier – as suitable – lagged values of the endogenous variable for subsequent first-differences as instruments, provided that the ε_{it} components of the errors are uncorrelated. Arellano and Bond (1991) have developed the framework for this "Difference GMM" estimator, which makes use of the maximum number of lags of the endogenous variable as instruments at each point in time. This is the estimator that we will use to model economic performance in (1).

Furthermore, Blundell and Bond (1998) show that under certain circumstances, a new estimator constructed by adding equations in levels to the equations in differences, for which suitable first-differences of the endogenous lagged variable are used as instruments, could bring significant improvements at the cost of only an additional assumption. This estimator, the System GMM estimator, works much better especially if the series are close to being a random walk, so that their first-differences are close to being innovations, or if the variance of the permanent effects (u_i) is large relative to the variance of the transitory shocks (ε_{it}).

4. Data

Our data set consists of an average annual sample of 405 non-financial large publicly traded companies listed on the main international stock exchanges, leading to 2,027 firm-year observations from 2002 to 2006. [7] The sample covers nine industries defined according to the MSCI global industry classification standards (GICS): oil and gas (10.2%), basic resources, industrials (23.8%), consumer goods (16.9%), healthcare (6.5%), consumer services (13.6%), telecommunications (4.4%), utilities (9.9%), technology (8%), and basic materials (6.7%). We exclude financials since they have a role of intermediaries in the economy and their balance sheet structure is different from that of other sectors (which may adversely distort the distribution of the financial variables). Also, their CSR characteristics are structurally different than those of other sectors.[8] The regional distribution of the sample looks as follows: 45% European, 30% North-American, 15% Japanese, and 10% other.

Moreover, we also use an alternative rearrangement of the GICS sectoral classification in order to isolate several consumer-oriented sectors, as done by Siegel and Vitaliano (2007). Thus, in addition to basic resources (including the oil & gas producers and basic materials sectors), industrial goods, and industrial services (including oil & gas distributors), we consider consumer search goods, durable experience goods, non-durable experience goods, and experience services. For an exact classification, see Appendix B. This sectoral classification was only used to construct an alternative DEA-based CSR index to that based on a 10 industry classification.

The explanatory variables of interest are the two CSR aggregated indices, a DEA-based one and an equally-weighted average one, defined in Section 2.3. Both indices are constructed based on eight selected CSR dimensions that were rated every year of the analyzed period by Sustainable Asset Management, an asset management company specialized in sustainable investments. These CSR dimensions are: codes of conduct/bribery & corruption, corporate citizenship, corporate governance, eco-efficiency, human capital, risk management, talent attraction, and social reporting. Two other available CSR dimensions were excluded from the analysis due to their high correlation with the remaining ones. These are: environmental reporting (showing a 0.61 correlation with eco-efficiency and a 0.55 correlation with social reporting) and labor practices (>0.5 correlation with both labor practices and codes of conduct).

[7] The sample was reduced from an initial 2,092 non-financial sample for which CSR data was available as follows: 17 observations lost to missing values, 27 lost to negative book value of equity, 1 lost to Tobin's Q larger than 10, 9 lost to ROA or lagged ROA smaller than -25%, 4 lost to ROA or lagged ROA larger than 40%, 2 lost to capital intensity larger than 1.5, and 5 lost to change in sales larger than 1.05.

[8] Some authors (e.g., Siegel and Vitaliano, 2007; Baron et al., 2009), however, do not exclude Financials.

As mentioned, the dependent variable is Return on Assets (ROA) and Tobin's Q.[9] These two, together with the other control variables described in Section 3, were obtained based on data from Worldscope Datastream. Eight industry dummies, four year dummies, and four regional dummies were also included among the explanatory variables.

No wave of data is lost when using lagged ROA among the explanatory variables, as lagged ROA is saved as a separate variable and recorded even for the first year of data. One wave of data is however lost for Tobin's Q, and therefore the sample size reduces in this case to 1,372 observations. By estimating Model (1) in first-differences, an additional wave of data is lost in both cases.

5. Empirical results

5.1 Descriptive statistics and properties of our DEA-based CSR index

We begin our analysis by taking a look at how our DEA-based index behaves when calculated within industry, using the industry classification defined in accordance with Siegel and Vitaliano (2007). Summary statistics of the original DEA efficiency ratios are provided in Table 1. Recall that these DEA ratios are sample-specific and therefore comparisons of the ratios *per se* are not meaningful between samples. Also, a DEA ratio equal to one indicates an efficient firm, and the higher the ratio, the higher the degree of inefficiency. A typical DEA ratio sample distribution would be skewed to the right, indicating that most firms are on the frontier or close to it with fewer firms the further away from the frontier one goes.

The industry-wise DEA ratio statistics (Table 1) indicate that most firms in the consumer search goods and basic resources sectors are *relatively* efficient, as their average DEA ratios are closest to 1. Firms in these two sectors seem to behave more homogeneously with respect to CSR than those in the other sectors, as they achieve comparable values in all eight CSR dimensions. At the other end, firms in the experience services sector behave heterogeneously with respect to CSR, with only a small proportion (37%) being CSR efficient relative to the others.

[9] See the Appendix for additional data details.

Table 1. Summary Statistics of DEA ratio by industry

	Obs.	Mean	St. Dev.	Min.	Max.	% efficient
Basic resources	316	1.04	0.09	1	1.80	63%
Industrial goods	352	1.07	0.11	1	1.50	46%
Industrial services	154	1.05	0.08	1	1.37	54%
Cons. search goods	69	1.02	0.06	1	1.29	75%
Durable exp. goods	450	1.05	0.08	1	1.41	56%
Non-Durable exp. goods	281	1.06	0.10	1	1.66	50%
Experience services	405	1.08	0.11	1	1.67	37%

Operating the transformation of the industry-wise DEA efficiency score into a three-value discrete variable (2 for CSR efficiency, 1 for slight inefficiency, and 0 for high inefficiency) enables comparison of this CSR index between firms in different industries. A first look at the values in Table 2, Column 1, suggests that within each industry, a majority of firms are either efficient or slightly inefficient, as their mean values range from 1.05 in experience services (the industry with the most inefficient firms) to 1.67 in consumer search goods (the industry with the least inefficient firms). Unfortunately, this industry-wise index cannot offer information about which industries do more than others in terms of strategic CSR. Siegel and Vitaliano (2007) claim that in sectors that provide products with characteristics that can only be assessed after purchase (such as the experience goods sector), CSR could be a useful tool in reducing asymmetric information between producers and consumers. These sectors are therefore more likely to embrace CSR. In order to find out which industries are more CSR strategic than others, we calculated a CSR index at the regional level (Table 2, Column 2). By calculating its mean by sectors, we can determine which sector contains the highest concentration of CSR-strategic firms (the highest mean value). The results are intriguing.

The most CSR-efficient firms, by a large margin, are found in the basic resources sector with a mean CSR index of 1.50, followed by the non-durable experience goods sector with a mean CSR index of 1.13[10] (Table 2, Column 2). This might sound surprising, but it could be explained by the fact that such firms face increased pressure and scrutiny with respect to CSR (or may even face tougher regulation, which cannot be disentangled based on our raw CSR data) and therefore achieve more in this respect; or they simply operate in a business area that gives them a lot more scope for improving their CSR standing than firms in other industries.

[10] The difference is statistically significant

Table 2. Summary statistics by industry for the CSR index-dummy built by industry (Column 1**)** and region (Column 2**)** and for the average CSR index (Column 3).

		(1)		(2)		(3)	
		CSR industry-wise		**CSR region-wise**		**Average CSR**	
	Obs.	**Mean**	**St. Dev**	**Mean**	**St. Dev**	**Mean**	**St. Dev**
Cons. search goods	69	**1.67**	0.63	**1.04**	0.86	60	10
Basic resources	316	**1.51**	0.71	**1.50**	0.70	69	11
Industrial services	154	**1.39**	0.73	**1.01**	0.82	56	11
Durable exp. goods	450	**1.37**	0.79	**1.26**	0.78	65	10
Non-Durable exp. goods	281	**1.27**	0.81	**1.13**	0.81	61	13
Industrial goods	352	**1.19**	0.84	**1.01**	0.85	57	11
Experience services	405	**1.05**	0.83	**0.98**	0.83	58	13

The CSR index is a dummy variable taking the value 2 for efficient firms (DEA ratio=1), 1 for slightly inefficient firms (1<DEA ratio<=1.1), and 0 for substantially inefficient firms (DEA ratio>1.1). Average CSR is the average CSR score across the eight CSR dimensions.

In line with findings in Siegel and Vitaliano (2007), the durable and non-durable experience goods sectors have a higher proportion of firms that behave CRS strategically than any other consumer-oriented sector. A contrasting finding is that the experience services sector, also consumer-oriented, shows the lowest proportion of strategic CSR firms. The mean CSR index is actually slightly biased toward highly CSR-inefficient firms (with a mean value below 1 in Table 2, Column 2, which is statistically different from 1.26 – the mean value in durable experience goods). One explanation could be that their result, which was especially strong for the financial services subsectors and much weaker for the other services subsectors, cannot be replicated here as the financial sector was excluded. This raises a caution flag when dealing with CSR and financials in general.

In Table 2, we have also presented summary statistics for a simple average CSR index (in Column 3), which puts equal weights on the CSR dimensions instead of the endogenously determined weights in the DEA-based index. When comparing it with the between-industries regional DEA-CSR index (Column 2), one can observe that they generally deliver an identical ranking of industries in terms of CSR standing. The mean value differences of either index between any two of the sectors, i.e, industrial goods, industrial services and experience services, are not statistically significant at any conventional level. This correspondence between the two very different indices indicates that, generally, firms that are best-in-sample along one particular

CSR dimension tend to do well along other dimensions as well. Thus, extreme cases (and by some undesirable) such as firms that are terrible on many dimensions (e.g., spilling large amounts of oil into the sea, workers dying as a result of work accidents, etc.) but really good at one (which could be "human capital" here) are rare or non-existent.

At the same time, the two indices do not convey an identical message everywhere. For example, while the average CSR index indicates that firms in the non-durable experience sector do statistically more with respect to CSR than firms in the industrial services sectors, the DEA-CSR index indicates that they are in fact comparable (the difference in DEA-CSR index mean values is not statistically significant at conventional levels).

Looking at these findings differently, we could infer that our constructed CSR index indicates, once again, that CSR can also serve other purposes than reducing information asymmetry between producers and consumers, as it does not materialize strongly in all consumer experience goods or services sectors and it is even strongly present in industries not dealing directly with consumers (like basic resources).

5.2. Econometric analysis

A unilateral comparison of our DEA-CSR index does not reveal all its features. In order to evaluate it in greater detail, we also investigated how it relates to economic performance and, more importantly, whether its relationship to economic performance differs from that of an average CSR index. To this end, we estimated Model 1 for ROA and Tobin's Q, using alternatively our DEA-based index and an average index as the explanatory variable of interest. For the econometric analysis, the DEA index is calculated industry-wise, based on a 10 industry classification. It is expected that by measuring a firm's CSR relative to its industry peers, defined in a strict and not broad sense, one can better capture any strategic elements of CSR, and thus it increases the chances of capturing any existing link with economic performance through the econometric analysis.

The summary statistics for the 2,027 firm-year observations in Table 3 indicate that the average firm has a 6.63% ROA and a 1.63 Tobin's Q and that it is close to being CSR efficient (CSR index 1.32). Also, most firms are large – with an average size of $11.8 million and the others tightly distributed around this average. Moreover, the average firm has a 20% level of long-term debt to assets and an average life of more than 23 years.

Regarding the strategic and average CSR indices, it is noteworthy that our strategic CSR index has a five times higher variation than the average CSR index. Multi-collinearity (untabulated results) among the explanatory variables in Model 1 is not a concern. Most coefficients lie between the reasonable values of -.28 (between firm size and lagged Tobin's q) and .26 (between dividend-to-book and lagged Tobin's q). One can also note that while our strategic CSR index has a .25 correlation coefficient with firm size, an average CSR index has a correlation coefficient of .42 – i.e., almost twice as high. Thus, our CSR measure reduces the typically found bias of higher CSR performance among larger companies. Also, the correlation between our CSR measure and the average CSR index is .64.

Table 3. Summary statistics for all explanatory variables in Model (1)

Variable	**Obs.**	**Mean**	**St. Dev.**	**Min**	**Max**
ROA (%)	2027	6.63	5.76	-22.31	37.18
Tobin's Q	2027	1.34	0.92	0.12	8.99
Strategic CSR	2027	1.32	0.80	0	2
Average CSR	2027	61.45	12.31	19.49	94.22
Roa_t-1 (%)	2027	6.06	5.68	-22.72	36.66
Tobin's Q_t-1	1376	1.31	0.91	0.12	6.52
ln(TobinsQ_t-1)	1376	0.10	-0.57	2.13	1.87
ln(tobinsQ)	2027	0.12	-0.56	2.13	2.20
Firm Size (ln)	2027	16.28	1.26	11.30	20.44
Firm Risk	2027	0.20	0.13	0.00	0.85
Sales growth	2027	0.07	0.12	-0.48	0.69
Cap. Intensity	2027	0.19	0.12	0.00	1.31
Div-to-Book	2021	0.07	0.15	0.00	3.31
Age (years)	2027	23.59	11.19	0.48	42.94

By modeling the cross-sectional and time variation in the financial and strategic CSR performance of firms through Model 1, one can note the positive relation between strategic CSR and either return on assets or firm value (Tables 4 and 5, Column 1).[11] After controlling for industry, region, and year-specific effects as well as lagged ROA and the other controls, we observe a 0.22% higher ROA for a firm that goes either from highly inefficient to slightly

[11] All estimations were done using Stata10.1.

inefficient, or from slightly inefficient to efficient. A similar, though statistically weaker, positive relation can be observed between strategic CSR and firm value (Table 5, Column 1). However, as soon as firm-specific effects are taken into account (Tables 4 and 5, Columns 3), any such effect vanishes and we find evidence of no impact of strategic CSR on ROA or Tobin's Q. This holds even after fully correcting for the endogenous nature of lagged ROA – due to correlation with the fixed effects component of the error term – as the consistent estimates obtained using the Arellano-Bond Difference GMM technique still show no relationship between our strategic CSR measure and ROA and Tobin's Q respectively (Tables 4 and 5, Column 5).[12] Finally, it is noteworthy that all models behave well in the sense that the effects of the control variables have the sign predicted by the established theory.

A few words regarding the legitimacy of relying on the Difference GMM estimator for our sample are required. Fixed effects are important in our panel data, as they account for more than 70% of the variation of the compound error terms (.73 for ROA and .99 for Tobin's Q). At the same time, there is no evidence either from the Hansen test or from the serial correlation tests (m2 and m3) that either of these AR(1) models are misspecified and thus we can trust the instruments we have used. Moreover, both ROA and Tobin's Q autoregressive coefficient estimates (0.14 and 0.80 respectively) lie within the ranges given by the OLS and Fixed Effects estimates (0.09-0.63 for ROA and 0.32-0.88 for Tobin's Q), which is reassuring. However, we were still led to believe that, in the model for Tobin's Q, both consistency and efficiency could be improved by using a System GMM estimator, i.e., by exploiting the correlations between first-differences lagged dependent and its level. The indications came from the high autoregressive coefficient for Tobin's Q, .80, corroborated with a very large variance of permanent shocks (a 99% of total error term variance due to fixed effects). In spite of this, such an estimator was not validated, as the extra-overidentifying restrictions – of zero correlation between the first-differences lagged dependent variable (instrument) and the compound error term in levels – did not hold (Difference-in-Hansen test of 0.02). This could be due to the additional assumption about initial conditions of System GMM, i.e., that the fastest growing individuals (i.e., firms) in terms of Tobin's Q are not systematically closer or farther from their steady-states than slower-growing ones (Roadman, 2006), not holding in our sample. Thus, the difference GMM estimator performs best on our sample(s).

[12] The Difference GMM was implemented as a one-step estimator with Windmeijer-corrected cluster-robust errors.

Table 4. Pooled OLS (Col. 1, 2); Fixed-Effects (Col. 3, 4); Difference GMM (Col. 5, 6).

Estimation of Model (1) for ROA, with either a strategic CSR measure (Columns I) or an average CSR measure (Columns II) as explanatory variable.

Variables	(1) OLS (I)	(2) OLS (II)	(3) FE (I)	(4) FE (II)	(5) Diff-GMM (I)	(6) Diff-GMM (II)
Strategic CSR	**0.22****		**-0.07**		**0.00**	
	(0.05)		**(0.64)**		**(0.99)**	
Average CSR		**0.02****		**-0.01**		**-0.04****
		(0.05)		**(0.45)**		**(0.03)**
Lag(ROA)	0.63***	0.63***	0.09*	0.09*	0.14**	0.14**
	(0.00)	(0.00)	(0.07)	(0.06)	(0.04)	(0.04)
Firm Size	-0.27***	-0.29***	-2.74***	-2.75***	-2.10**	-2.21**
	(0.00)	(0.00)	(0.01)	(0.01)	(0.02)	(0.01)
Firm Risk	-3.26***	-3.24***	-10.53***	-10.58***	-11.41***	-11.61***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Sales Growth	0.78	0.83	2.96*	2.99*	3.01*	3.08*
	(0.41)	(0.38)	(0.06)	(0.06)	(0.06)	(0.06)
Capital Intensity	1.72	1.72	5.86***	5.86***	3.40**	3.33**
	(0.12)	(0.12)	(0.01)	(0.01)	(0.04)	(0.04)
Constant	6.76***	6.48***	49.39***	50.32***		
	(0.00)	(0.00)	(0.00)	(0.00)		
Year Dummies	Y	Y	Y	Y	Y	Y
Industry Dummies	Y	Y	N	N	N	N
Region Dummies	Y	Y	N	N	N	N
m2	-	-	-	-	0.67	0.67
m3	-	-	-	-	0.47	0.47
Rho(FE)			0.73	0.73		
Hansen	-	-	-	-	0.49	0.50
# Overid. restr.	-	-	-	-	10	10
Observations	2027	2027	2027	2027	1376	1376
Groups	-	-	615	615	464	464
R-sq.	0.55	0.55	0.18	0.18	-	-
F-statistic	74.10	76.34	16.22	16.31	14.74	14.86
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

P-values in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Rho indicates the fraction of total variance due to that of fixed effects u_i. m2, and m3 are second-order and third-order serial correlation in the first-differenced residuals, asymptotically N(0,1). Hansen is a test of the validity of overidentifying restrictions for the GMM estimators. # Overid. restr. indicates number of overidentifying restrictions tested by Hansen test. P-values are reported for all tests.

Table 5. Pooled OLS (Col. 1, 2); Fixed-Effects (Col. 3, 4); Difference GMM (Col. 5, 6). Estimation of Model (1) for log (Tobin's Q), with either a strategic CSR measure (Columns I) or an average CSR measure (Columns II) as explanatory variable.

	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	FE	FE	Diff-GMM	Diff-GMM
Variables	(I)	(II)	(I)	(II)	(I)	(II)
Strategic CSR	**0.01***		**-0.00**		**0.00**	
	(0.09)		**(0.73)**		**(0.84)**	
Average CSR		**0.00***		**-0.00**		**-0.00**
		(0.07)		**(0.35)**		**(0.51)**
Lag Tobin'sQ	0.88***	0.88***	0.32***	0.32***	0.80***	0.82***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Firm Size	-0.03***	-0.03***	-0.28***	-0.28***	-0.33***	-0.33***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Firm Risk	-0.14***	-0.14***	-0.02	-0.02	-0.09	-0.09
	(0.00)	(0.00)	(0.85)	(0.84)	(0.48)	(0.47)
Sales Growth	-0.06	-0.06	-0.06	-0.06	-0.07	-0.07
	(0.11)	(0.12)	(0.41)	(0.42)	(0.45)	(0.43)
Capital Intensity	-0.03	-0.03	-0.03	-0.03	-0.20***	-0.21***
	(0.49)	(0.49)	(0.70)	(0.71)	(0.01)	(0.01)
DivToBook	0.13***	0.13***	0.11**	0.11**	0.08***	0.08***
	(0.01)	(0.01)	(0.04)	(0.03)	(0.01)	(0.01)
age (Years)	0.00	0.00	0.08	0.08	0.34**	0.34**
	(0.50)	(0.51)	(0.50)	(0.49)	(0.02)	(0.02)
Constant	0.64***	0.62***	3.03	3.05	-	-
	(0.00)	(0.00)	(0.32)	(0.31)		
Year Dummies	Y	Y	Y	Y	Y	Y
Industry Dummies	Y	Y	N	N	N	N
Region Dummies	Y	Y	N	N	N	N
m2	-	-	-	-	0.43	0.43
Rho(FE)	-	-	0.98	0.98	-	-
Hansen	-	-	-	-	0.82	0.82
# Overid. restr.	-	-	-	-	5	5
Observations	1372	1372	1372	1372	906	906
Groups	-	-	464	464	377	377
R-sq.	0.91	0.91	0.34	0.34	-	-
F-statistic	541.92	537.19	28.22	28.29	29.36	28.80
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

P-values in parentheses; *** p<0.01, ** p<0.05, * p<0.1; Rho indicates the fraction of total variance due to that of fixed effects u_i; m2 is the second-order serial correlation in the first-differenced residuals, asymptotically N(0,1). Hansen is a test of the validity of overidentifying restrictions for the GMM estimators. # Overid. restr. indicates number of overidentifying restrictions tested by Hansen test. P-values are reported for all tests.

As noted by Baron et al. (2009), the difference between the estimates with and without firm fixed effects (i.e., Difference GMM vs. OLS) could be due to the situations of the firms prior to inclusion in the data sample. For some reason, some firms could have had both high strategic CSR and high economic performance, whereas others could have had low strategic CSR and low economic performance. These relations could then have persisted over the period of the analysis, which would explain the positive and significant coefficient of strategic CSR in the OLS estimation. These findings emphasize once again the necessity to control for unobserved firm heterogeneity when modeling CSR and firm performance since when this is not controlled for, inexistent positive or negative relationships between the two might emerge.

The behavior of the estimated impact of the average CSR index across the various estimations closely mimics that of strategic CSR, with one substantial distinction (Columns II versus Columns I, Tables 4 and 5). Its estimated impact on return on assets, though *initially positive*, *turns negative* when controlling for firm-specific effects (Column 6, Table 4). This finding actually emphasizes some of the benefits of our DEA-based index, as outlined in Section 2.1. Based on the same underlying CSR data, it becomes clear now that one aggregation technique might provide a CSR index with a persistent negative impact on firm performance, while another technique might reveal a neutral impact, leaving the analyst with the difficult task of confronting the conflicting results, when, in fact, the relationship between CSR and firm performance ultimately is dictated by the type of CSR one is measuring. Thus, the preference for one technique or another will depend on the kind of CSR one would like to measure. As claimed in Section 2.3, our DEA-based index can capture more of a *strategic* feature of CSR while a simple average of the same underlying CSR scores could measure *other non-strategic* component of CSR, e.g., altruistic or social pressure CSR. This alleged dichotomy, in aggregation terms, is further emphasized by the distinct impacts they have on firm performance, the latter negative and the former neutral. As indicated by Baron (2001), undertaking CSR for altruistic reasons or in response to a threat by an activity might lead to lower profitability as it is only implies additional costs.

Unlike the predictions of Baron (2001), however, our data analysis did not find evidence for a positive association between our strategic CSR measure and profitability; rather, a neutral association. Yet, the relationship is not negative, indicating that strategic CSR might in fact be undertaken only up to the extent that marginal costs equal marginal benefits. In a further analysis on regions, we found some evidence that the lack of an overall effect might be due to contrasting effects on different regions, with a positive effect in Continental Europe and a negative in Anglo-Saxon countries, yet these results were not statistically significant. It could be that a richer dataset in each region, with a larger sample in each industry and a more detailed industry classification, would be helpful in revealing stronger effects. However, this is left for future research.

Our results are related to several lines of empirical findings on the link between CSR performance and economic performance. First, not only do we investigate the impact of environmental performance on profitability measures as most studies in this area (Russo and Fouts, 1997; Waddock and Graves, 1997; Hart and Ahuja, 1996), we also account for other CSR dimensions, as only a few other studies do (e.g., Manescu and Starica, 2007). Second, we show that firm-specific effects and past economic performance – which is likely to drive CSR performance to some extent in the next period – play an important role in explaining the relationship between economic performance and CSR. Finally, we have shown that *strategic* CSR has a neutral impact on both return on assets and firm value, while *an equally-weighted index* of the same CSR dimensions has a negative impact on return on assets.

5.3 Robustness checks

Our results are robust to a variety of alternatives used either for building the DEA-based index or measuring Tobin's Q. First, we tried different alternatives for the transformation of the DEA ratios into the discrete-variable CSR index (e.g., a two-value dummy variable, bundling together the slightly and highly inefficient firms; extending the efficient CSR firm to include firms that are very close to the efficient frontier – with DEA ratios lower or equal to 1.05; or a three-value dummy variable, similar to the one currently used, but with more or less restrictive cutoff points than 1.1) with overall qualitatively similar results, statistically stronger in some cases and weaker in others.

Then, a CSR index built on a 7 sectors versus a 10 industry classification was also used in the econometric analysis, yielding a stronger impact on ln(Tobin's Q), but a lower impact on return on assets (in the pooled OLS estimation). Third, different alternatives of calculating

Tobin's Q as in Guenster el al. (2005) or Elsayed and Paton (2009) do not qualitatively alter the results. Finally, using Tobin's Q directly, instead of its natural logarithm transform, produces statistically weaker results.

6. Conclusions

In empirical investigations where a quantitative measure of corporate social responsibility is required, one of the main difficulties is to account for the multidimensional and heterogeneous nature of the concept. It is difficult to aggregate company achievements with respect to various CSR dimensions in a way that leads to a fair and meaningful index.

This paper proposes a novel method based on Data Envelopment Analysis (DEA), a mathematical model traditionally used for efficiency analyses, to aggregate various CSR dimensions while considering the notion of strategic CSR, as proposed by Baron (2001) and argued for in Porter and Kramer (2006). We assume that most managers correctly identify and favor dimensions of CSR that might provide their companies with competitive advantages, and our constructed CSR index accounts for this strategic CSR behavior.

Statistical properties of this DEA-based CSR index, calculated at the regional level and based on a set of eight dimension-specific CSR scores provided by SAM, a specialized rating agency, indicate that most CSR-strategic firms, by a large margin, are in the basic resources sector. This might be explained either by the fact that firms in this sector are faced with high environmental pressures and therefore have to deliver more in this respect, or by the fact that this sector offers a larger scope for engagement in CSR. Additionally, our strategic-CSR measure provides support for the claim that strategic CSR may be used to reduce asymmetric information between producers and buyers (Siegel and Vitaliano, 2007), as firms in the durable and non-durable experience goods sectors are found to behave CSR strategically to a larger extent than in any other consumer-oriented sector.

Furthermore, we also tested whether strategic CSR is profit-enhancing, as expected in theory, since it is motivated by profit-maximization. Employing a Difference-GMM estimation technique, in a dynamic framework, we analyzed the relationship between our strategic CSR index and economic performance, measured by ROA and Tobin's Q, using a 5-year panel of 405 non-financial publicly traded large corporations. Our findings indicate that firm-specific effects are found responsible for a positive estimated relationship between economic performance (measured by Return on Assets, ROA, and Tobin's Q) and both our measure of strategic CSR, built at the industry level (on a ten industry classification), and an equally-weighted CSR index. Once these effects are controlled for, we find a neutral relationship between strategic CSR and economic performance, and a negative relationship between ROA and the equally-weighted CSR index. It could be that, for some reason, before entering the panel, firms have both high (low) economic performance and high (low) levels of CSR, and that this persists throughout the analysis period, which emphasizes the need to control for firm-specific effects. The contrasting estimated impacts on profitability of the two CSR indices, which were based on the same underlying CSR indicators but aggregated differently, is evidence that aggregation matters greatly when investigating CSR and firm performance. It also suggests, once again, that our DEA-based CSR index might succeed in capturing strategic CSR.

While additional research is needed to explain the diverse reasons (e.g., social pressure, profit-maximization, altruism) as to why firms adopt a CSR stance (Siegel and Vitaliano, 2007), the evidence presented here supports the view that strategic CSR can be compatible with a profit-maximizing strategy as it does not hurt economic performance.

References:

Anderson, T. W. and C. Hsiao, 1981, "Estimation of dynamic models with error components," *Journal of the American Statistical Association*, 76, 598-606.

Arellano, M. and S. R. Bond, 1991, "Estimation of dynamic models with error components," *Review of Economic Studies*, 58, 277-297

Bagnoli, M. and S. Watts, 2003, "Selling to socially responsible consumers: competition and the private provision of public goods," *Journal of Economics & Management Strategy*, 12(3), 419–445.

Banker, R.D., A. Charnes, W.W. Cooper, (1984) "Some models for estimating technical and scale inefficiencies in data envelopment analysis," *Management Science*, 30, 1078-1092.

Baron, D. P., 2001, "Private politics, corporate social responsibility, and integrated strategy," *Journal of Economics & Management Strategy*, 10(1), 7-45.

Baron, D. P., 2009, "A positive theory of moral management, social pressure, and corporate social performance," *Journal of Economics and Management Strategy*, 18(1), 7-43

Baron, D. P., M.A. Harjoto and H. Jo, 2009, "The Economics and politics of corporate social performance," *Stanford University Graduate School of Business Research Paper No. 1993; Rock Center for Corporate Governance at Stanford University Working Paper No. 45. Available at SSRN: http://ssrn.com/abstract=1202390*

Belu, C., 2009, "Ranking corporations in terms of sustainable and socially responsible practices," *Sustainable Development*, 17(4), 257-268.

Besley, T. and M. Ghatak, 2007, "Retailing public goods: The economics of corporate social responsibility," *Journal of Public Economics,* 91, 1645-1663.

Blundell, R. W. and S. R. Bond, 1998, "Initial conditions and moment restrictions in dynamic panel data models," *Journal of Econometrics*, 87, 115-143.

Bond, S., 2002, "Dynamic panel data models: a guide to microdata methods and practice," *CeMMAP working papers CWP09/02*, Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Burke L. and J. M. Logsdon, 1996, "How Corporate Social Responsibility Pays Off," *Long Range Planning*, 19(4), 495-502.

Charnes A., W. W. Cooper and E. Rhodes, 1978, "Measuring efficiency of decision making units," *European Journal of Operational Research*, 2, 429–444.

Cooper, W. W., L. M. Seiford and K. Tone, 2000, *Data Envelopment Analysis*, Kluwer, USA.

Crane A., A. Mcwilliams, D. Matten, J. Moon, D.S. Siegel, 2008, *The Oxford Handbook of Corporate Social Responsibility*, Oxford University Press, New York, USA.

Elsayed, K. and D. Paton, 2005, “The Impact of Environmental Performance on Financial Performance: Static and Dynamic Panel Data Evidence,” *Structural Change and Economic Dynamics*, 16, 395–412.

European Commission, 2001, The Directorate General for Enterprise and Industry, http://ec.europa.eu/enterprise/csr/index_en.htm [31 December 2008].

Fama, E. F. and K. R. French, 2000, “Forecasting profitability and earnings,” *Journal of Business*, 73(2), 161-175.

Färe, R., S. Grosskopf, C.A.K. Lovell , 1989, ”Multilateral productivity comparisons when some outputs are undesirable: a non-parametric approach,” *Review of Economics and Statistics*, 71, 90-98.

Guenster N., R. Bauer, J. Derwall, and K. Koedijk, (2005). “The economic value of corporate eco-efficiency,” *European Financial Management*, forthcoming, [doi: 10.1111/j.1468-036X.2009.00532.x]

Hart, S. and G. Ahuja, 1996, “Does it pay to be green? An empirical examination of the relationship between emission reduction and firm performance,” *Business Strategy and the Environment*, 5, 30-37.

Heslin, P. A. and J. D. Ochoa, 2008, “Understanding and developing strategic corporate social responsibility,” *Organizational Dynamics*, 37(2), 125-144.

King, A. and M. Lenox, 2001, “Does it really pay to be green? An empirical study of firm environmental and financial performance,” *Journal of Industrial Ecology*, 5(1), 105-116.

Lovell, C. A. K. and J. T. Pastor, 1997, “Target setting: An application to a bank branch network,” *European Journal of Operational Research*, 98, 290-299.

Lovell, C. A. K. and J. T. Pastor, 1999, “Radial DEA models without inputs or without outputs,” *European Journal of Operational Research*, 118, 46-51.

Manescu, C. and C. Starica, 2007, “Do Corporate Social Responsibility scores predict firm profitability? A case study on the publishers of the Dow Jones Sustainability Indexes,” *Mimeo*, Dept. of Economics, University of Gothenburg.

Margolis, J., H. Elfenbein, and Walsh, J., 2007, “Does it pay to be good? A meta-analysis and redirection of research on the relationship between corporate social and financial performance,” *Working paper*, Available at:

http://stakeholder.bu.edu/2007/Docs/Walsh,%20Jim%20Does%20It%20Pay%20to%20Be%20Good.pdf [Accessed 10.09.2010].

Miller, M. and F. Modigliani, 1961. “Dividend Policy, Growth, and the Valuation of Shares,” *The Journal of Business*, 34, 411-433.

Porter, M. E. and M. R. Kramer, 2006, “Strategy and Society: The Link Between Competitive Advantage and Corporate Social Responsibility,” *Harvard Business Review*, December, 78-92.

Porter, M. and C. Van der Linde, 1995, “Green and competitive: ending the stalemate,” *Harvard Business Review*, Sept-Oct, 120-134.

Russo, M. V. and P. A. Fouts, 1997, “A resource-based perspective on corporate environmental performance and profitability,” *Academy of Management Journal*, 40(3), 534-559.

Siegel, D. S. and D. F. Vitaliano, 2007, “An empirical analysis of the strategic use of Corporate Social Responsibility,” *Journal of Economics & Management Strategy*, 16(3), 773-792.

Waddock, S. A., and B. S. Graves, 1997, “The Corporate Social Performance-Financial Link,” *Strategic Management Journal*, 18(4), 303-319.

United Nations Environment Programme-Finance Initiative and Mercer Joint Report, 2007, “Demystifying Responsible Investment Performance: A review of key academic and broker research on ESG factors”

APPENDIX A. **Data description**

Compilation of CSR data is usually a tedious process. For example, CSR performance in year *t* is reflected in the CSR scores published in September of year *t+1*. We therefore need to ensure that both the dependent variable (i.e., return on assets) and the explanatory variables (i.e., CSR variable and controls) are *contemporaneous*. As the ends of the fiscal years of the firms in our samples range from January to December, we have designed a representative matching rule according to which there is at least a six month overlap between the period reflected by the CSR measure and the financial data. Thus, CSR data for year *t* (which in fact reflects CSR performance in year *t*-1) is linked to financial data for either January-May in year *t* or for June-December in year *t*-1, depending on the firms' fiscal year ends.

APPENDIX B. **Industrial classification**

The regrouping of the Datastream sectoral classification, in order to obtain the sector classification used in Siegel and Vitaliano (2007), is as follows:

I. Basic resources sector includes oil & gas producers (Datastream Industry code 530) and basic materials (1000).

II. Industrial goods sector includes construction materials (2300), aerospace & defense (2710), general industrials (2720), electronic and electric equipment (2730), and industrial engineering (2750).

III. Industrial services sector includes oil equipment, services and distribution (570), industrial transportation (2770), and support services (2790).

IV. Consumer search goods sector includes personal goods (3760) and furnishings (3726)

V. Durable experience goods sector includes automobiles & parts (3300), durable household products (3722), home constructions (3728), leisure goods (3740), pharmaceutical & biotech (4570), medical equipment (4535), and utilities (7000).

VI. Non-durable experience goods sector includes technology (9000), food & beverages (3500), non-durable household products (3724), and tobacco (3780).

VII. Experience services sector includes consumer services (5000), telecommunications (7000), healthcare providers (4533), and healthcare suppliers (4537).

Paper III

Stock returns in relation to environmental, social, and governance performance: Mispricing or compensation for risk?[ab]

Cristiana Mănescu [c]

Abstract

Using detailed data on seven environmental, social, and governance (ESG) attributes for a long panel of large publicly-traded U.S. firms during July 1992-June 2008, only community relations were found to have had a positive effect on risk-adjusted stock returns, which effect was not compensation for risk but could be due to mispricing. Additionally, a changing effect of employee relations was found from positive during July 1992-June 2003 to negative during July 2003-June 2008. The positive effect could be due to mispricing, whereas there is some evidence that the negative effect was compensation for low non-sustainability risk. A weak negative effect of human-rights and product safety indicators on risk-adjusted stock returns in the more recent period was also found to be likely due to mispricing. The implications are that certain ESG attributes might be value relevant but they are not efficiently incorporated into stock prices.

JEL: C22, G12, G14, M14

Keywords: SRI, sustainability, risk-factor test, market efficiency

[a]*This article is forthcoming under the same title during Spring 2011 in a Special Issue on Responsible Investments, of the Sustainable Development Journal, published by John Wiley & Sons Ltd.*

[b]*This research was conducted as part of the research programme Behavioral Impediments to Sustainable Investment, funded by the Swedish Foundation for Strategic Environmental Research, MISTRA.*

[c]Centre for Finance, School of Business, Economics and Law, University of Gothenburg, Box 600, SE-40530, Gothenburg, Sweden. Tel.: +46-317864412. Fax: +46-317861326. Email: cristiana.manescu@cff.gu.se I am grateful to Bert Scholtens (Editor) and an anonymous referee, to my supervisors Cătălin Stărica and Martin Holmen, and also to Leonardo Becchetti (discussant at Licentiate Seminar), Hossein Asgharian (discussant at Final Seminar), Evert Carlsson, Stefano Herzel, Lennart Hjalmarsson, Gordon Clark, Nils Kok, Eric Knight, Stephanie Giamporcaro, Constantin Belu, Rick Wicks, and Andreea Mitrut for valued input; as well as to seminar participants: at the 15th Annual International Sustainable Development Research Conference; at the Mistra Workshop Gothenburg; at the Oikos PRI Young Scholars Finance Academy; at the XVII International Tor Vergata Conference on Banking and Finance; at the GU Department of Economics; and at GU's Centre for Finance. I appreciate the use of financial data from Keneth French's website. Special thanks to Lisa Symonds of KLD Research & Analytics for quick answers regarding data questions. The usual disclaimer applies.

1. INTRODUCTION

Incorporating environmental, social, or governance (ESG) concerns into investment practices - so-called socially responsible investing - is increasingly popular (SIF, 2007). One of the major forces behind the surge in such investment is increasing demand for it from institutional and even individual investors, partly because of increasing awareness of environmental risk (e.g., climate change) and social risk (e.g., the crisis in Sudan) (SIF, 2007, :iv).

The trade-off between ESG performance and investment returns is difficult to analyze, both theoretically and empirically, primarily because of the multi-dimensionality of the ESG concept. On the one hand, most empirical evidence suggests that "good" stocks, i.e., with high ESG scores, earn positive abnormal returns (Derwall et al., 2005; Statman and Glushkov, 2009). This is claimed to be due either to investors underestimating the benefits of ESG or overestimating its costs, i.e., mispricing the value relevance of ESG concerns, or to compensation for risk (Derwall et al., 2005). At the same time, few studies show that some good stocks earn negative abnormal returns, also explained as either mispricing or compensation for risk (Derwall and Verwijmeren, 2007).

On the other hand, there is strong evidence that "shunned" stocks, i.e., of companies in the alcohol, gambling, tobacco, firearms, military, and nuclear industries, also earn positive abnormal returns (Statman and Glushkov, 2009; Hong and Kacperczyk, 2009). This is explained as the effect of social norms as norm-constrained investors discriminate against those companies, thus producing a "neglect" premium in their risk-adjusted returns (Hong and Kacperczyk, 2009). A theoretical model developed by Heinkel et al. (2001) shows that, with a sufficient number of norm-constrained investors the lack of risk-sharing opportunities in the market will lead to higher cost of equity capital (i.e., higher expected returns) for shunned stocks (i.e., polluting firms in their model) and a lower cost of equity capital for acceptable stocks (i.e., non-polluting firms in their model).

The controversy about what relative risk-adjusted returns should be expected from ESG investments arises because of the non-exclusiveness of the two stock universes. "Good" firms are generally those with an outstanding record with respect to at least one of the several ESG concerns. At the same time, "shunned" firms are not necessarily those with the lowest record with respect to actual ESG scores, but are rather ignored by ethical investors simply because of the industry they are in, i.e., because of ethical beliefs.

This study analyzes whether it is mispricing or compensation for risk that underlies the positive or negative abnormal returns earned by firms with high ESG scores. First we will evaluate if there is an association between those ESG scores and the cross-section of stock returns, while controlling for other factors known to explain stock returns, such as beta, size, value, and momentum. Then, for those ESG concerns for which we find an effect we will implement the Charoenrook and Conrad (2005) test of whether the candidate ESG concern is a risk factor. If the required conditions for the test are not met, we can conclude that the effect observed is due to mispricing.

The ESG scores used here are based on the Kinder, Lydenberg, and Domini (KLD) Research and Analytics ratings of U.S. firms (1991-2006), the longest existing dataset on ESG concerns including community relations, corporate governance, diversity, employee relations, environment, product safety, and human rights dimensions. Two alternative aggregation techniques are used, together with a Best-in-Class method, which is helpful in addressing industry variation in the ESG scores. As large companies are alleged to disclose more ESG information, which is thus readily available for investors' decision-making, the focus here is on components of either the Standard and Poor's 500 index (S&P500) or the Domini Social 400 index (DS400). Using the components of these two indexes also minimizes the presence of companies for which a score of zero is likely to indicate lack of rating as opposed to neutral ESG performance.

Fama and MacBeth (1973) month-by-month cross-sectional regressions of monthly stock returns on the beta, size, book-to-market, and momentum risk-factors, as well as on seven ESG concerns (community relations, corporate governance, diversity, employee relations, environment, human rights, and product safety) found that only the community relations indicator had a positive impact on stock returns. The risk-factor test shows, however, that the higher returns were not compensation for risk, but might reflect mispricing.

In a further analysis on subperiods July 1992-June 2003 and July 2003-June 2008, we will find that there is a shift in the effect of some ESG concerns, consistent across several aggregations of ESG scores. During July 1992-June 2003, community relations and employee relations had a positive effect on stock returns, while during July 2003-June 2008 employee relations, human rights, and product safety had a negative effect. The risk-factor test finds some evidence of employee relations as a risk factor in the second period, but not for any of the others. Thus, firms with poor employee relations seem to carry a premium for high non-sustainability risk. The evidence is weak as only the necessary condition for employee relations is met, but not the sufficient condition. One explanation for the shift in the effect of employee relations is that investors may now more efficiently value good employee relations in the recent period, following an increase in the availability of such information. On the other hand, the effects of all other factors are found to be likely due to mispricing.

The remainder of the paper is organized as follows: section 2 discusses the theoretical framework on the expected returns of socially responsible investments and sets the working hypotheses; section 3 presents the empirical strategy; section 4 describes how the ESG scores and the financial variables are obtained as well as how the risk-factor mimicking portfolios are constructed; section 5 presents and discusses the empirical results, and section 6 summarizes and draws conclusions.

2. Theoretical framework and hypotheses

Socially responsible investing (SRI) emerged in the U.S. in the 1960s, some decades later in the UK, Canada, and Australia (Kinder, 2005). Because it mainly involved aligning investors' portfolios with their values, it has been called **values-based SRI** (Kinder, 2005). By the late

1990s it had also developed another name, **value-seeking SRI**, which means identifying and incorporating into the investment process social and environmental criteria which might affect future financial performance (Kinder, 2005). Thus, SRI is an evolving notion and it is still unclear *why* investors would incorporate *any* of the ESG concerns in their investment process, as well as *which* ESG concerns they might prioritize.

Academics have hypothesized various explanations of high-ESG firms' expected returns relative to those of conventional firms mainly based on two arguments, one economic and one based on discriminatory tastes (Statman, 2006). The *economic argument* is that there are costs and benefits associated with ESG concerns, but it is not clear which are higher. Moreover, in order for costs and benefits to be efficiently reflected in stock prices, there must be sufficient information available on ESG performance for the market to incorporate it efficiently. According to *the discriminatory-tastes argument*, however, the relation between costs and benefits of ESG is secondary. Instead there might be enough investors who derive non-financial utility from ESG investing that they affect pricing irrespective of whether ESG is net costly or beneficial. Furthermore, there is a third argument which has attracted less attention, a *non-sustainability risk* argument. ESG performance might in fact affect the risk profile of firms by adding a non-sustainability risk component in addition to the market risk, size, book-to-market and other systemic risks documented theoretically and empirically.

Putting these arguments together, and their implications for stock returns, leads to three mutually-exclusive scenarios regarding the risk-adjusted returns of high-ESG firms relative to low-ESG firms, which are briefly discussed below. Renneboog et al. (2008) provide a critical review of the literature on SRI and thoroughly discuss the causes and the shareholder-value impacts of ESG, among other related issues.

"The no-effect scenario" is that there is no difference in the returns, adjusted for common risk-factors, of high-ESG firms relative to low-ESG firms. This is entirely consistent with the efficient markets hypothesis if the ESG performance of firms provides no information relevant for pricing (Statman and Glushkov, 2009). Even if ESG performance provides information relevant for pricing, if this information is publicly available and fully incorporated into asset prices, then there should still be no difference in the risk-adjusted returns of ESG and non-ESG firms (Wall, 1995). In this case we cannot distinguish whether ESG costs are higher or lower than ESG benefits by looking solely at stock returns. When controlling for common risk-factors in studies on socially responsible (SR) fund performance versus non-SR fund performance, this scenario has usually been confirmed (Bauer et al., 2005).

"Mispricing scenario" is that ESG performance has an impact on firms' cash-flow streams (i.e., it is value relevant[1]) but without enough information available it is not efficiently reflected in stock prices. This translates into either higher or lower risk-adjusted returns for high-ESG firms depending on the net benefit of ESG (by the economic argument). If the benefits of ESG outweigh

[1]This is the *value relevance hypothesis* discussed in Derwall and Verwijmeren (2007)

their costs, but investors, on average, consistently underestimate the benefits or overestimate the costs, then the risk-adjusted returns of high-ESG firms would be higher than of low-ESG firms (Statman, 2006). Underestimating the benefits of ESG could be reflected in positive earnings surprises (Edmans, 2008) or in reduced earnings volatility (Derwall and Verwijmeren, 2007), either of which could lead to mispricing. High performance with respect to the environment during 1995-2003 (Derwall et al., 2005) or employee relations during 1984-2005 (Edmans, 2008) have been found to provide positive abnormal returns, interpreted as mispricing. In a sample of multinationals, Dowell et al. (2000) also found that firms with high environmental standards had higher firm value than others, measured by Tobin's Q. Portfolios built on specific ESG dimensions have even been shown to have positive abnormal returns over long periods (Kempf and Osthoff, 2007).

On the other hand, it could be that, by the same economic argument, the risk-adjusted returns of high-ESG firms will be lower if the benefits of ESG are lower than its costs and incompletely informed investors overestimate benefits or underestimate costs. Barnea and Rubin (2006) show that ESG performance can be a source of agency costs because firms' managers have an incentive to promote ESG investment at the expense of shareholders' benefits in order to gain reputational benefits. Their hypothesis has empirical support as they found a negative link between insiders' ownership and the social rating of firms.

"Risk-factor scenario" is that expected returns of low-ESG firms are higher primarily because they carry a premium for non-sustainability risk. The ESG rating of a company might indicate its exposure to a non-sustainability risk factor. Besides environmental risk, this risk factor might include product and commercial-practices risks, or risk associated with workplace quality of life (Dufresne and Savaria, 2004). It can also relate to litigation risk, investor trust, and other intangible advantages (Becchetti and Ciciretti, 2006; Derwall and Verwijmeren, 2007) that might dramatically affect firms' future financial performance (i.e., value-seeking SRI). Given the increased awareness of sustainability risk, it is expected that the non-sustainability premium has increased in recent years.

A systematic distortion in market pricing resulting in higher expected returns for low-ESG firms (or non-SR firms) can *also* occur under the discriminatory-tastes argument that divestment from certain stocks is done exclusively on moral or ethical grounds (Hong and Kacperczyk, 2009; Derwall and Verwijmeren, 2007). The existence of investors who derive non-financial utility from investing with high-ESG firms (i.e., values-based SRI), could decrease the demand for the stock of low-ESG firms and thus increase their cost of equity capital. Heinkel et al. (2001) developed an equilibrium model in which there is an impact on firms' costs of capital as soon as there is a substantial share of SR investors.[2] Though expressed in environmental terms, the model could

[2]This model belongs to the family of theoretical models explaining empirical pricing anomalies (e.g., the equity-size premium anomaly) through differences in information quality or quantity (asymmetric information) across investors (Arbel et al., 1983; Merton, 1987). These models show that in equilibrium it is both market risk and idiosyncratic risk that affect expected returns, due to distortions induced by different investor bases.

easily be extended to social and governance concerns. Hong and Kacperczyk (2009) have shown that higher institutional divestment from ”sin stocks” leads to higher cost of capital for these firms relative to others, confirming this alternative argument.

Alternatively, under the risk-factor scenario, it can also be that expected returns of high-ESG firms are higher because they might carry a premium for some missing risk-factors, others than the common factors beta, size, value and momentum. This argument has usually been used as alternative explanation to mispricing for the higher risk-adjusted returns observed for high-ESG firms, such as the eco-efficiency premium puzzle (Derwall et al., 2005).

Either because of ethical beliefs (values) or a (non-)sustainability risk-factor, ESG performance might thus affect expected stock returns, if two conditions are met: Information on ESG performance must be available to investors, and there must be enough investors who care. Since both of these can change, the effect of ESG performance on stock returns can vary over time.

Regarding availability of ESG information, the UK Companies Act of 2006 holds company directors responsible for disclosing environmental and social information regarding firms' long-term prospects (Clark and Knight, 2009). According to CorporateRegister.com, a private company specializing in tracking corporate social responsibility/ESG disclosures, since the mid-1990s more and more firms have started disclosing ESG information (Dhaliwal et al., 2009). Dhaliwal et al. argue that the increasing voluntary ESG disclosures might have to do with increased scrutiny of the corporate impact on society following the loss of trust after the series of corporate scandals in the early 2000s. Finally, KPMG (2008) reports that 80% of the 2200 largest corporations worldwide published an ESG report in 2008 or integrated ESG information into their annual reports.

Regarding the number of socially responsible investors, Haigh and Hazelton (2004) question the existence of a large number, while Hong and Kacperczyk (2009) argue the opposite. Epstein and Freedman (1994) surveyed a random sample of average individual investors and found that they also wanted ESG information, not just institutional investors. Those surveyed primarily wanted information on environmental performance and product quality. Empirical studies also show market reaction to disclosure of such information (Patten, 1990), confirming that investors care.

The working hypothesis here is that, during 1992-2008, there was a positive association between candidate ESG variables and risk-adjusted stock returns, due to ESG benefits outweighting costs, but without enough ESG information available. This hypothesis will be confirmed if the Charoenrook and Conrad (2005) test rejects the null that any particular ESG variable is a risk-factor. Moreover, given the rise in sustainability-risk awareness and in information availability in recent years, we can also expect that, for subperiods, some of the ESG variables might function as systematic risk-factors (i.e., non-sustainability risk-factors). Therefore, the second working hypothesis here is that in recent years we might find a negative association between some ESG variables and risk-adjusted stock returns which could be due to compensation for lower sustainability risk. Again, this will be tested by the Charoenrook and Conrad (2005) test.

Many studies have pointed towards such positive or negative associations but without explaining what lies behind them. Edmans (2008) is among the very few who tested for mispricing (by looking at positive earnings surprises) as explanation for the positive association between employee relations and risk-adjusted stock returns.

This study aims at narrowing this gap by explicitly testing whether any risk-adjusted returns reaction to ESG performance is compensation for risk or rather is due to mispricing. It may also help to better understand the puzzle that one observes both high returns for high-ESG firms and high returns for ethically excluded firms, formulated by Statman and Glushkov (2009).

3. The Empirical Strategy

The central aim of this paper is to test whether the explanatory power that some ESG concerns have for stock returns is due to mispricing or compensation for risk.

Explanatory power is assessed using cross-sectional regression of excess stock returns on ESG scores and four factors known to explain stock returns: beta, size, value, and momentum, i.e., the Fama-French three-factor model (Fama and French, 1992) plus the momentum factor identified by Jegadeesh and Titman (1993). The cross-sectional approach is preferred to the portfolio approach of analyzing only the return differential between high- and low-ESG stock portfolios because of interest in the monotonic effect of ESG concerns on stock returns. Moreover, the use of ex-post returns should supplement Derwall and Verwijmeren (2007), who used measures of ex-ante returns, i.e., implied cost of equity capital.

The test for the effect of ESG concerns is whether their estimated effect is statistically indistinguishable from zero as the four-factor model predicts. Alternatively, if ESG concerns are relevant for explaining variation in returns, we can implement the Charoenrook and Conrad (2005) test of whether a candidate variable is a risk factor or not. The test is whether there is a relationship between the conditional mean and the conditional variance of the return of the candidate variable's factor-mimicking portfolio, and whether risk, in the mean-variance framework, explains all the return of a mimicking portfolio.

The central economic question that the cross-section regressions can answer is why average returns vary across assets (Cochrane, 2005). The expected return of an asset should be high if it has large exposure to factors that carry risk premiums, i.e., market risk or beta, size (market capitalization), book-to-market ratio, or momentum (a simple average of 10 past returns).

The four-factor model estimated here is extended with several ESG variables - or alternatively with an aggregate ESG variable - as

$$R_{jt+1} = \gamma_0^{t+1} + \gamma_1^{t+1}\widehat{\beta_{jt}} + \gamma_2^{t+1} Size_{jt} + \gamma_3^{t+1} BookToMarket_{jt} + \gamma_4^{t+1} Momentum_{jt} + \gamma_5^{t+1} ESG_{jt} + \theta_{jt+1} \tag{1}$$

where the excess stock return for firm j in month $t+1$ (R_{jt+1}) is a function of $\widehat{\beta_{jt}}$, the estimated market risk (beta) of the firm; $Size_{jt}$, the firm's log of market capitalization; $BookToMarket_{jt}$, its book-to-market ratio; $Momentum_{jt}$, the average return over the period $t-2$ and $t-12$ months; ESG_{jt}, seven individual firm ESG variables or an overall ESG variable; and an i.i.d. error term, θ_{jt}, with zero mean and constant variance. Size and book-to-market ratio are updated monthly (as did Galema et al., 2008), while estimated beta and the ESG variable(s) are updated every year. Beta is estimated for each asset j ($j = 1 ... N$) through a time-series regression up to time t of the asset's returns and the market-index return.

Empirical evidence shows that not all ESG concerns are equally relevant to investors or, perhaps more importantly, that there are confounding effects between some of them. Thus the focus on the individual ESG concerns in the analysis, also complemented with an aggregate ESG score.

This model (1) was estimated with the Fama-MacBeth procedure, which allows for time-varying coefficients ($\widehat{\gamma_k^{t+1}}$, $k = 1, 2, 3, 4$). This is known as a *two-pass* Fama-MacBeth estimation, because one first estimates a time-series regression of individual stock returns on the market-index return in order to obtain beta estimates, and then uses these estimates as explanatory variables in the cross-sectional regression (1). Estimation bias can be a problem when using this procedure, due to possible measurement error (sampling variance) in $\widehat{\beta}_j$, i.e., the errors-in-variables (EIV) problem (Black et al., 1972). The estimated $\widehat{\beta}_j$ in (1) equals the true (unobservable) value plus a measurement error (or sample variance) v_j, assumed to be i.i.d. with zero mean and variance σ_v^2. The "grouping technique" developed by Black et al. (1972) provides N-consistent estimates of $\widehat{\gamma}_1$ and was thus used in what follows (see detailed description of this approach in Appendix A).

Fama and MacBeth (1973) indicate that the mean values of the $\widehat{\gamma_k^{t+1}}$ coefficients, which are needed for statistical inference, can be computed as the time series averages of the estimated coefficients, i.e., $\bar{\gamma}_k = Avg(\widehat{\gamma}_k) = \sum_{t=1}^{T} \widehat{\gamma}_k^t / T$. Then the t-test is the ratio of this mean and the time-series standard error of the estimated coefficient, i.e., its time-series standard deviation ($sd(\widehat{\gamma}_k)$) multiplied by the square root of the time-series length (T), $t(\widehat{\gamma}_k) = \frac{\bar{\gamma}_k}{sd(\widehat{\gamma}_k)/\sqrt{T}}$, $k = 0, 1, 2, 3, 4$).

Most empirical studies of ESG point out that ESG is industry-specific. For example, firms in a certain industry might have both high ESG scores and high stock returns, while those in another industry might have low ESG scores and low returns. Without controlling for such industry effects, a false positive association might thus appear between ESG and returns. Conversely, any of the ESG concerns that might have different effects across industries would obscure their overall effect. Therefore, Equation (1) was also estimated augmented with 9 industry dummies Ind_i, to control for any confounding effects, as

$$R_{jt+1} = \gamma_0^{t+1} + \Sigma_{i=1}^{9} \alpha_i^{t+1} Ind_i + \gamma_1^{t+1} \widehat{\beta_{jt}} + \gamma_2^{t+1} Size_{jt} + \gamma_3^{t+1} BookToMarket_{jt} + \\ + \gamma_4^{t+1} Momentum_{jt} + \gamma_5^{t+1} ESG_{jt} + \theta_{jt+1} \quad (2)$$

which is the main estimation model.

The test for a candidate ESG variable being a risk factor was developed by Charoenrook and Conrad (2005) based on the relationship between the conditional mean and the conditional variance of return on the candidate variables's factor-mimicking portfolio. Specifically, with a linear factor structure, they note that the risk premium at time t on the factor-mimicking portfolio, like that of any asset, is related to the conditional variance of the portfolio's return with the pricing kernel. However, for a factor-mimicking portfolio, the conditional variance with the pricing kernel is linearly related to the factor's conditional variance, and hence to the portfolio's variance. Therefore, there must be a linear relationship between the conditional mean and the conditional variance of the factor-mimicking portfolio return, if that factor is a component of the pricing kernel. For a formal deduction of this relation see Section III in Charoenrook and Conrad (2005).

To test this relation, they propose a Garch-in-Mean family of estimations, among which the main one is

$$\begin{aligned} R_{X,t+1} &= \mu + \delta h_{t+1} + \eta_{t+1} \\ with \qquad h_{t+1} &= \omega + \alpha \eta_t^2 + \gamma h_t \end{aligned} \quad (3)$$

where $R_{X,t+1}$ is the excess return on the mimicking portfolio for factor X at time $t+1$; $h_{t+1} = Var_t(R_{X,t+1})$ is the expected conditional variance for the factor portfolio at time $t+1$; and $\eta_{t+1}/\sqrt{h_{t+1}}$ is normally distributed.

Three hypotheses were tested related to Equation (3). First, if X is a risk factor, then the relation between the conditional mean and variance of the portfolio return, estimated with coefficient δ in Equation (3), should have the same sign as the conditional expected risk premium on X (and be statistically significant).

Second, the intercept term (μ) in the conditional mean equation of (3) should be zero, meaning that the expected risk premium for the portfolio should be given entirely by the portfolio's conditional variance. A non-zero and statistically significant intercept might represent a component of the factor-mimicking portfolio unrelated to risk or an unmodeled time-variation in the price of (variance) risk.

Third, if X is a proxy for risk, then the Sharpe ratio of the factor-mimicking portfolio should be plausible, i.e., it should be less than that of the *ex ante* tangency portfolio. MacKinley (1995) argues that a reasonable value for this is approximately 0.17 per month, while Ahn et al. (2005) found a 0.39 Sharpe ratio from an efficient frontier built with 25 size- and book-to-market sorted portfolios for the 1959-2003 sample period.

Here, Equation (3) was estimated for mimicking portfolios for those ESG concerns found to explain excess returns in the cross-sectional approach. If the conditions above are not met, we have evidence that those ESG concerns' effects on stock returns were due to mispricing or irrational investor behavior. If all conditions are met, we have evidence that the ESG concerns functioned as risk-factors.

4. The Data

4.1. **The ESG Dataset.** Measuring ESG concerns is not easy, and measuring it over a long period as required for risk-factor analysis is especially difficult. Therefore the expertise of the investment research firm Kinder, Lydenberg, and Domini (KLD) Research & Analytics was used. They are the leading authority on social research and indexes for institutional investors (their database is preferred in many empirical analyses on SRI). Since 1991 they have collected data about seven ESG dimensions: community relations, corporate governance (labeled Other until 2002), diversity, employee relations, environment, human rights, and product safety; for 650 publicly-traded U.S. firms (components of either the S&P 500 or the DS400 index) until 2000, then 1100 firms during 2001-2002, and about 3100 (components of the Russell 3000 index) since 2003. The data comes from media articles, company documents, and direct communication with company officers, as well as government and NGO information. The research process, however, is proprietary. The previous year's annual ESG dataset is available for purchase at the end of each January or early February.

Each ESG dimension is graded annually based on a number of positive and negative indicators, i.e., strengths and weaknesses, which are given a score of 1 if present, otherwise 0. As KLD data has been used extensively in the ESG literature, the complete list of indicators can be found in Appendix B (for a thorough description of the data see Becchetti and Ciciretti, 2006; Derwall and Verwijmeren, 2007). The total number of indicators has varied from 54 in 1991 to 72 in 2006. The strengths and weaknesses for each ESG dimension are then aggregated to give an overall score for that dimension.

The most popular aggregation method, used in most studies based on KLD data (e.g., Derwall and Verwijmeren, 2007; Chatterji et al., 2008), has been to take the sum of strengths net of the sum of weaknesses, i.e.,

$$ESG_t^j = \sum_{s=1}^{u_t^j} strength_s^j - \sum_{r=1}^{k_t^j} weakness_r^j \tag{4}$$

where ESG_t^j = ESG dimension j, year t; u_t^j = number of strengths for dimension j, year t; $strength_s^j$ = strength indicator, equal to 1 if the firm meets strength s, j, otherwise 0; k_t^j = number of weaknesses for dimension j, year t; $weakness_r^j$ = weakness indicator, equal to 1 if the firm meets weakness r, j, otherwise 0.

The drawback with this approach is lack of comparability across years and across dimensions, as the number of strength and weakness indicators for most have varied. The product safety and environment dimensions have been quite stable since 1991, but the other dimensions have varied considerably.

An alternative aggregation method, used by Kempf and Osthoff (2007) (hereinafter the ”KO” method), involves first converting weaknesses into strengths by taking binary complements (meaning that if a certain weakness is not present, i.e., rated as 0, it is considered a strength rated with 1; if the weakness is present, then its corresponding strength is rated 0) and then summing up over all ”strengths” for a certain ESG concern. The results are then normalized to maximum for each dimension, by

$$ESG_t^j = \frac{\sum_{s=1}^{u_t^j} strength_s^j + \sum_{r=1}^{k_t^j} (!weakness_r^j)}{u_t^j + k_t^j} \tag{5}$$

where $!weakness_r^j$ = transformed strength indicator, equal to 1 if the firm does not meet weakness$_{r,j}$, otherwise 0; and ESG_t^j, u_t^j , k_t^j , and $strength_s^j$ are the same as in Equation (4). The scores thus range from 0 to 1.

This method is still subject to two criticisms concerning the representativeness and firm-level comparability of the results for at least five of the seven ESG dimensions.

The first criticism mainly concerns the environmental dimension, which is especially relevant in industries such as oil & gas, basic materials, industrials, and healthcare, while less relevant in others. Thus, for example, if the weakness ”ozone depleting chemicals” (as well as other weaknesses in the environmental dimension) are not present, this can certainly be seen as a strength for the oil & gas industry, but it is irrelevant for the telecom and financial industries. Without adjustment for this, firms in industries less affected by environmental strengths or weaknesses would receive inflated environment scores. Similarly a weakness indicator in the community dimension is ”investment controversies”, defined as ”a financial institution (!) that had lending or investment practices that lead to controversies”. Again, without adjustment, firms in other industries would receive inflated community scores.

The other criticism of the KO method is that not having certain weaknesses (such as having recently paid substantial fines for violations of employment standards) automatically becomes a ”health and safety” strength which competes with the already defined ”health and safety” strength (if the firm has made significant effort to improve employment conditions). However, not having paid fines for violations of employment standards is not the same as making effort for improvement. The firm has then complied with common rules, making it neutral with respect to the employee dimension, but not necessarily a leader. Examples in other ESG dimensions are ”South Africa” weakness in the human rights dimension, meaning that the company had faced controversies because of its operations in South Africa (which is irrelevant if the firm has no operations in

S.A., yet automatically gets a ”South Africa” strength), or ”indigenous people relations” weakness in the community relations dimension.

Thus the KO method might *systematically* distort the aggregate scores for at least four of the ESG dimensions: community relations, employee relations, environment, and human rights.

As both aggregation methods, i.e., Equations (4) and (5), are flawed, a new one answering these criticisms (hereinafter called the ”Relative” method) was constructed. It is based on Equation (4), but in addition standardizes the numbers of strengths and weaknesses in each dimension by the corresponding annual numbers of strengths and weaknesses (similar to the KO method). This avoids the two criticisms of the KO method and provides year-to-year comparability that was missing in the procedure in Equation (4). Formally, the method is

$$ESG_t^j = \frac{\sum_{s=1}^{u_t^j} strength_s^j}{u_t^j} - \frac{\sum_{r=1}^{k_t^j} weakness_r^j}{k_t^j} \tag{6}$$

By construction, ESG scores based on this method range from -1 to 1. The annual overall firm-ESG score across all seven dimensions is then simply their average ($\sum_{j=1}^{7} ESG_t^j/7$).

For comparison purposes, both the Relative and the KO method are used in the empirical analysis, but the Relative method is considered more suitable, not least because it is only a slight modification of the standard aggregation method used in almost all studies on KLD data.

In addition, a Best-in-Class score, by either the Relative or the KO method, is also used in order to embed industry differences already in the ESG scores. The Best-in-Class score is computed as the firm-score deviation relative to its industry mean score. This is a very frequent method in the ESG literature (e.g., Kempf and Osthoff, 2007; Statman and Glushkov, 2009). In the end, four measures for ESG performance are used: the Relative score, the KO score, the Relative Best-in-Class (Relative BC), and KO Best-in-Class (KO BC).

4.2. **The Financial Dataset.** The financial variables matched with the KLD data were obtained from Datastream. KLD data contains firm-year observations, identified since 1995 by their CUSIP code.[3] Missing CUSIPs before that were supplied by referencing firm-name and ticker identification with firms in 1995. Based on these CUSIPs, ISINs were next obtained, followed by Datastream codes, the unique security identification key in Datastream.[4] Datastream codes could be identified for 96.6% of the firms in the KLD dataset.

To ensure that accounting variables (in financial statements) were available before the stock returns they are used to explain, equity book-values for all fiscal year-ends in calendar year $t-1$ (1991-2006) were matched with returns for July of year t (1992-2007) to June of year $t+1$ (1993-2008). The 6-month (minimum) gap between fiscal year-end and return-period is conservative, similar to that used in Fama and French (1992).

[3]CUSIP (Committee on Uniform Security Identification Procedures) is the North American security identification code.

[4]ISIN=International Securities Identifying Number.

Market equities at the end of December of year $t-1$ were used to compute book-to-market ratios, while logs of market equities at the end of June of year t gave firm size. Both variables were updated for each month. Therefore, to be included in the return tests, a firm must have had available the book value (Item 03480 in Datastream) as of December 31 of year $t-1$ and market value of equity (Item MV) as of June 30 of year t. Moreover, in order to obtain the pre-ranking beta estimates, monthly returns for at least 24 months preceding July of year t must have also been available.

Monthly stock returns were obtained based on the Datastream Item Return Index.[5] The excess stock-return series used in the empirical tests is the stock return less the risk-free rate. A proxy for the monthly risk-free rate, as well as for the return series of U.S. market index, of SMB and HML portfolios, were obtained from K. French's website.[6]

The most sensitive variable in the empirical tests is the firm β estimate, which was obtained by applying the portfolio grouping technique as in Fama and French (1992) with a minor modification due to different sample sizes, described in Appendix A.

4.3. **The Risk-Factor Mimicking Portfolios.** The non-sustainability risk-factor mimicking portfolios mirror the book-to-market mimicking portfolio used in Fama and French (1993). Firms were first sorted independently in two size (Small, Big), two book-to-market (Value, Growth), and three ESG-score (sustainability) portfolios, resulting in 12 portfolios. The size and book-to-market breakpoints were the median sample market-equity and median book-to-market, respectively. The sustainability breakpoints which determined the Low, Medium, and High Sustainability portfolios are the 30th and 70th sample percentiles. The returns on the portfolios were value-weighted. (Equally-weighted portfolios were also used without altering the final results). The Low-Sustainability minus High-Sustainability (LMH) risk factor was defined as

$$\begin{aligned} LMH &= 1/4(SmallValueLowSust. + SmallGrowthLowSust. + BigValueLowSust. \\ &+ BigGrowthLowSust.) - 1/4(SmallValueHighSust. + SmallGrowthHighSust. \\ &+ BigValueHighSust. + BigGrowthHighSust.) \end{aligned}$$

Portfolio returns were computed monthly, from July of year t to June of year $t+1$, and rebalanced every year based on market equities as of June of year t, book-to-market as of December of year $t-1$, and ESG scores available Jan-Feb of year t. Using a ranking on size and book-to-market minimizes the effects of these factors on the LMH risk-factor of interest. This is the non-sustainability risk-factor mimicking portfolio on which the Charoenrook and Conrad (2005) test was later applied.

[5]The Item Return Index accounts for stock splits and dividends by adding back the dividend amount to the ex-dividend-date closing price.
[6]$http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html\#Benchmarks$, accessed September 2008.

5. Empirical Analysis

5.1. **Sample Description.** As noted, the sample consisted of components of the S&P500 and the DS400 index. The initial dataset consisted of an average of 639 firms per year for the 16 years from 1991 to 2006, with 10,220 firm-year observations, reduced to 10,174 that also had industry classification. Summary statistics for the seven individual ESG dimensions and an aggregate score, computed with the Relative and KO aggregation methods as well as their Best-in-Class equivalents, are presented in Table 1.

Panel 1: Relative Aggregation method (between -1 and 1)

	Community Rel.	Corp. Gov.	Diversity	Empl. Rel.	Env.	Human Rights	Prod. Safety	Aggreg. ESG
Mean	0.05	-0.10	0.04	0.02	-0.02	-0.03	-0.05	-0.02
SD	0.15	0.20	0.21	0.18	0.15	0.13	0.20	0.08
p25	0.00	-0.33	0.00	-0.03	0.00	0.00	0.00	-0.06
p50	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-0.01
p75	0.14	0.00	0.12	0.17	0.00	0.00	0.00	0.04

Panel 2: KO Aggregation method (between 0 and 1)

Mean	0.44	0.41	0.33	0.49	0.51	0.70	0.48	0.48
SD	0.09	0.11	0.12	0.10	0.08	0.17	0.10	0.05
p25	0.40	0.33	0.27	0.40	0.50	0.57	0.50	0.45
p50	0.40	0.43	0.27	0.50	0.50	0.71	0.50	0.48
p75	0.50	0.50	0.36	0.54	0.54	0.80	0.50	0.51

Panel 3: Relative Best-in-Class method

Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD	0.14	0.18	0.20	0.18	0.14	0.12	0.19	0.08
p25	-0.06	-0.11	-0.08	-0.08	-0.02	0.00	-0.05	-0.04
p50	-0.03	0.02	0.00	0.00	0.00	0.01	0.02	0.00
p75	0.06	0.10	0.12	0.12	0.04	0.04	0.08	0.05

Panel 4: KO Best-in-Class method

Mean	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SD	0.08	0.09	0.11	0.09	0.07	0.10	0.10	0.04
p25	-0.04	-0.05	-0.07	-0.04	0.00	0.00	-0.02	-0.02
p50	-0.02	0.01	-0.01	-0.01	0.00	0.01	0.01	0.00
p75	0.03	0.04	0.06	0.06	0.02	0.03	0.04	0.03

Table 1. Mean, standard deviation (SD), and 25th-, 50th-, and 75th-percentile values for community relations, corporate governance, diversity, employee relations, environment, human rights, and product safety scores and for an aggregate ESG score, calculated with the Relative and KO Aggregation as well as Best-in-Class methods, for the unbalanced panel of 10,174 firm-year observations during 1991-2006.

For both aggregation methods, higher score indicates better ESG performance. With respect to the Relative aggregation method (Table 1, Panel 1), a mean lower than 0 for corporate governance, environment, human rights and product safety dimensions indicates that the number of weaknesses outweighed the number of strengths. Identical values of 0 for the 25th-, 50th- and 75th-percentiles for environment, human rights and product safety dimensions show that 50% of firms recorded no strengths or weaknesses in these dimensions. As they are components of S&P500 or DS400, it is very unlikely that this is a result of KLD analysts not rating these firms at all.

While it is fairly easy to infer general characteristics of the ESG sample based on the Relative aggregation method, it is not straightforward to do so based on the KO method (Table 1, Panel 2). For example, a mean lower than 0.5 would indicate a low Relative ESG performance. However, by the KO measure, it might also indicate a much higher number of strengths than weaknesses, which imbalance tends to lower the aggregate scores. For example, diversity has a KO mean of 0.33, meaning low diversity performance at first, which is an artifact of the fact that there were 8 strengths and only 3 weaknesses measured. Standard deviations for most dimensions are about one-fifth of the corresponding mean, indicating low variation.

The Best-in-Class scores make the two aggregation methods directly comparable (Table 1, Panels 3 and 4). There are very similar distribution patterns of the two aggregation methods, except that both standard deviations and values of the Relative scores are double the size of the KO scores, which is achieved by construction. Estimated correlation coefficients between the ESG dimensions (not shown) are small, ranging from -0.15 to 0.20 regardless of method.

The requirement of minimum 24 monthly stock returns prior to July of year t (1992-2007) and equity book-value as of December 31 of year $t-1$ (1991-2006) reduced the sample to 9,371 firm-year observations. Eliminating outliers in monthly stock returns (over 60% a month) and in book-to-market ratios (negative and larger than 3) reduced the sample further to 9,202 firm-year observations, or 110,427 firm-month observations.

The logs of market equities (Table 2, Panel 1) indicate, as expected, that the sample consists of large firms, which is also indicated by an average beta lower than 1 and negative skewness in its distribution. The sample firms were classified into ten industry groups (sample distribution in parentheses): oil & gas (4.94%), basic materials (7.68%), industrials (18.95%), consumer services (14.02%), healthcare(7.39%), consumer goods (15.38%), telecom (2.09%), utilities (7.39%), financials (12.69%), and technology (9.47%).

5.2. **Analysis Results and Discussion.** Equations (1) and (2) were estimated using Fama-MacBeth month-by-month cross-sectional regressions, with the ESG variables computed with the Relative and KO methods along with their Best-in-Class counterparts. During 1992-2008, book-to-market and momentum explained the cross-section of returns while beta and size did not, irrespective of whether industry effects were controlled for or not (Tables 3 and 4). Moreover, as expected, book-to-market and momentum had positive effects on stock returns, the magnitudes of each, i.e., risk

Panel 1

	Mean	SD	p25	p50	p75
Excess Returns (%)	0.71	9.71	-4.39	0.65	5.73
Estimated Beta	0.91	0.36	0.61	0.87	1.08
Book-to-Market Ratio (%)	0.45	0.34	0.24	0.38	0.57
Market Equity (ln)	22.21	1.57	21.21	22.26	23.24
Momentum (%)	0.85	3.05	-0.74	0.85	2.42

Panel 2: Low-Minus-High Mimicking Portfolios (in %) 1992-2003

Community Relations	-0.10	1.54	-1.00	0.08	0.84
Employee Relations	-0.15	1.26	-0.97	-0.08	0.79
Human Rights	-0.33	2.55	-1.57	-0.50	0.91
Product Safety	0.11	2.17	-0.94	0.01	1.11

Panel 3: Low-Minus-High Mimicking Portfolios (in %) 2003-2008

Community Relations	-0.07	0.92	-0.69	-0.19	0.55
Employee Relations	0.38	1.22	-0.38	0.13	1.14
Human Rights	-0.22	1.62	-1.17	-0.14	0.69
Product Safety	0.20	1.21	-0.79	0.16	1.05

TABLE 2. Panel 1: Mean, standard deviation (SD), 25th-, 50th- and 75th-percentile values for monthly excess returns; estimated beta; book-to-market ratio; log of market equity; and momentum factor; for the unbalanced panel of 9,202 firm-year observations during 1991-2006. Panel 2 (3): Mean, standard deviation (SD), 25th-, 50th- and 75th-percentile values for community-relations, employee-relations, human-rights and product-safety monthly risk-factor mimicking portfolios (in %) during July 1992-June 2003 (July 2003-June 2008), where the ESG variables are calculated using Relative BC method.

premiums, being influenced by the units of measurement of the underlying factor, so they must be interpreted with caution. For example, with industry dummies, a one-standard-deviation increase in book-to-market ratio and in momentum led, respectively, to 1.39-1.22% and 2.56% higher average annual returns, based on the marginal estimated effects.

With respect to the ESG variables, only community relations had an effect on stock returns, which was especially strong when industry effects were controlled for (Table 4). In this case, community relations, computed by both Relative and KO methods as well as their Best-in-Class counterparts, had positive effect on stock returns, while without industry dummies (Table 3) a statistically significant effect is found only when using the Relative Best-in-Class method, which also controls to some extent for industry effects. Thus it is important to control for industry effects on ESG scores, and the Best-in-Class method appears to be the superior aggregation method. In terms of economic value, a one standard-deviation increase in the community relations score led to 0.52% (based on Relative scores) or 0.58% (based on KO scores) higher returns per year.

Without industry dummies	Relative		KO		Relative BC		KO BC	
Estimated Beta	0.08 (0.41)	0.07 (0.42)	0.08 (0.41)	0.07 (0.42)	0.07 (0.42)	0.06 (0.43)	0.07 (0.42)	0.05 (0.41)
Book-to-Market Ratio	0.42** (0.02)	0.38** (0.02)	0.43** (0.01)	0.39** (0.02)	0.44** (0.01)	0.41** (0.02)	0.44** (0.01)	0.41** (0.02)
Size	-0.01 (0.45)	-0.02 (0.34)	0.00 (0.47)	-0.02 (0.37)	0.00 (0.48)	-0.01 (0.40)	0.00 (0.48)	-0.01 (0.42)
Momentum	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)	0.08*** (0.00)
Aggregate ESG	0.03 (0.47)		0.03 (0.47)		0.22 (0.32)		0.46 (0.30)	
Community Relations		0.12 (0.29)		0.36 (0.17)		0.24* (0.10)		0.43 (0.11)
Corporate Governance		-0.05 (0.40)		-0.26 (0.26)		-0.01 (0.48)		-0.13 (0.36)
Diversity		-0.06 (0.36)		-0.25 (0.24)		-0.10 (0.26)		-0.36 (0.13)
Employee Relations		0.14 (0.19)		0.36 (0.12)		0.13 (0.20)		0.31 (0.15)
Environment		-0.10 (0.35)		-0.23 (0.34)		-0.12 (0.29)		-0.24 (0.30)
Human Rights		-0.46 (0.08)		-0.70 (0.11)		-0.24 (0.21)		-0.46 (0.17)
Product Safety		-0.09 (0.33)		-0.21 (0.30)		-0.06 (0.37)		-0.11 (0.36)
Intercept	0.53 (0.33)	0.73 (0.23)	0.45 (0.37)	0.85 (0.17)	0.45 (0.35)	0.58 (0.28)	0.45 (0.35)	0.55 (0.29)
Mean Adj. R^2 Cross-section	0.07	0.07	0.07	0.07	0.06	0.06	0.06	0.07
Max. Adj. R^2 Cross-section	0.37	0.40	0.37	0.38	0.37	0.38	0.37	0.37
N	639	639	639	639	639	639	639	639
T	192	192	192	192	192	192	192	192

P-values in parentheses. $*p <= 0.10$, $**p <= 0.05$, $***p <= 0.01$.

N=average number of cross-sectional observations. T=number of monthly observations.

Mean (Max) Adjusted R^2 is the average (maximum) adjusted R^2s of the T cross-sectional estimations.

TABLE 3. Fama-MacBeth estimation of Model (1) over July 1992-June 2008, where the ESG variables are computed using Relative, KO, Relative Best-in-Class, and KO Best-in-Class methods. The dependent variable, excess returns, is in percentage points (%).

The aggregate ESG variable had no statistically significant effect on stock returns over this period, which is evidence of confounding effects between the ESG dimensions (since the community relations score was significant) and therefore they should be analyzed individually. This result is consistent with the findings in Derwall and Verwijmeren (2007) for the lack of effect of an aggregate ESG variable.

With industry dummies	Relative		KO		Relative BC		KO BC	
Estimated Beta	-0.01 (0.47)	-0.02 (0.47)	-0.02 (0.47)	-0.03 (0.46)	-0.01 (0.47)	-0.02 (0.47)	-0.02 (0.47)	-0.03 (0.46)
Book-to-Market Ratio	0.34** (0.03)	0.30* (0.05)	0.34** (0.03)	0.30* (0.05)	0.34** (0.03)	0.30* (0.05)	0.34** (0.03)	0.30* (0.05)
Size	-0.02 (0.32)	-0.04 (0.19)	-0.02 (0.32)	-0.04 (0.20)	-0.02 (0.32)	-0.04 (0.19)	-0.02 (0.32)	-0.04 (0.20)
Momentum	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)	0.07*** (0.00)
Aggregate ESG	0.13 (0.38)		0.30 (0.36)		0.13 (0.38)		0.30 (0.36)	
Community Relations		0.29* (0.05)		0.54* (0.05)		0.29* (0.05)		0.54* (0.05)
Corporate Governance		-0.02 (0.46)		-0.15 (0.34)		-0.02 (0.46)		-0.15 (0.34)
Diversity		-0.04 (0.39)		-0.26 (0.20)		-0.04 (0.39)		-0.26 (0.20)
Employee Relations		0.11 (0.24)		0.27 (0.18)		0.11 (0.24)		0.27 (0.18)
Environment		-0.21 (0.17)		-0.41 (0.17)		-0.21 (0.17)		-0.41 (0.17)
Human Rights		-0.26 (0.19)		-0.50 (0.15)		-0.26 (0.19)		-0.50 (0.15)
Product Safety		-0.12 (0.23)		-0.24 (0.23)		-0.12 (0.23)		-0.24 (0.23)
Intercept	1.49* (0.10)	1.65** (0.05)	1.34 (0.15)	1.69** (0.05)	1.49* (0.10)	1.67** (0.05)	1.49* (0.10)	1.64** (0.05)
Mean Adj. R^2 Cross-section	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
Max. Adj. R^2 Cross-section	0.43	0.43	0.43	0.43	0.43	0.43	0.48	0.43
N	639	639	639	639	639	639	639	639
T	192	192	192	192	192	192	192	192

P-values in parentheses. *$p <= 0.10$, **$p <= 0.05$, ***$p <= 0.01$.

N=average number of cross-sectional observations. T=number of monthly observations.

Mean (Max) Adjusted R^2 is the average (maximum) adjusted R^2s of the T cross-sectional estimations.

TABLE 4. Fama-MacBeth estimation of Model (2) over July 1992-June 2008, where the ESG variables are computed using Relative, KO, Relative Best-in-Class, and KO Best-in-Class methods. The dependent variable, excess returns, is in percentage points (%).

In the model including industry dummies (Table 4), it makes no difference if one uses Best-in-Class ESG scores or plain scores, as the estimation produces identical results for each aggregation method and its Best-in-Class counterpart (Columns 1 and 3, or Columns 2 and 4). Industry dummies thus seem to completely capture the between-industry variation in the ESG scores. In addition, the Relative and KO scores deliver a very similar pattern of estimates, except for the obvious differences in magnitude.

Next, we investigate whether the positive effect of the community relations score was due to mispricing or whether it could be compensation for risk, with a premium for the firms with high community scores. For this reason a Low-Minus-High (LMH) community-risk-factor mimicking portfolio was built following the procedure described in Section 4.3, using the Relative aggregation method and its Best-in-Class counterpart. To ensure that the High and Low community portfolios are mutual exclusive, the 20th- and 80th-percentile breakpoints were used with the Relative scores, and the 30th- and 70th-percentile breakpoints were used with their Best-in-Class counterpart.

Garch-in-Mean estimation of Equation (3) for the community-mimicking portfolio, using either the Relative or the Relative Best-in-Class method of obtaining the scores, shows that the conditional mean return was not linked to its conditional variance, as the δ estimate was not statistically different from zero, i.e., p-value$_\delta$ $= 0.32$ or 0.40 (Table 6, third column). Thus, the positive effect of community relations identified in the cross-sectional approach (in Table 4) could not be compensation for risk. Moreover, the non-risk component of the mimicking portfolio, mu, had a negative sign (Table 6, third column), indicating that the unconditional mean return on the LMH community-portfolio could be negative. Though statistical significance is beyond conventional levels, as p-values were 0.22 or 0.21, its implication is consistent with that from the cross-sectional approach that there is a positive association between community relations scores and returns, which moreover could be due to mispricing. The benefits of good community relations might outweigh their costs, but stock prices might not properly discount this information since it is not publicly or widely available, which confirms the main working hypothesis and the mispricing scenario.

None of the other six ESG variables had any effect on stock returns, in either specification. By the no-effect scenario, this could be because the market values these ESG concerns efficiently, whether or not they carry any relevant information. However, a sign-shift in the effect of these ESG concerns on risk-adjusted returns during this period could also result in zero overall effect. Therefore, as robustness check of these findings, the effect of ESG performance on stock returns was estimated for two subperiods, 1992-2003 and 2003-2008. The 2003 breakpoint was chosen as there is reason to expect a change in ESG effect when more ESG information became available to investors. Table 5 reports estimation results for Equation (2) using the two plain ESG aggregation methods. The Best-in-Class method was dropped here as it has been shown that the inclusion of industry dummies yields similar estimates.

Human rights and product safety scores had consistent negative effects on stock returns in both periods, though weakly statistically significant only in the later period. Conversely, community relations had a consistent positive effect but statistically significant only in the earlier period. The estimated economic magnitude of the marginal effect for a one-standard-deviation increase in the score (Relative or KO) is substantial: +0.59% or +0.71% per annum for community relations in 1992-2003, -0.92% or -2.18% per annum for human rights in 2003-2008, and -0.77% or -0.71%

	1992-2003				2003-2008			
	Relative		KO		Relative		KO	
Estimated Beta	-0.06 (0.42)	-0.07 (0.41)	-0.07 (0.42)	-0.08 (0.40)	0.10 (0.37)	0.10 (0.37)	0.09 (0.37)	0.09 (0.38)
Book-to-Market Ratio	0.62*** (0.00)	0.59*** (0.00)	0.62*** (0.00)	0.59*** (0.00)	-0.27 (0.19)	-0.34 (0.13)	-0.26 (0.20)	-0.35 (0.12)
Size	0.00 (0.47)	-0.01 (0.44)	0.00 (0.48)	-0.01 (0.46)	-0.08 (0.12)	-0.11** (0.04)	-0.07 (0.13)	-0.11** (0.04)
Momentum	0.07** (0.01)	0.07** (0.01)	0.07** (0.01)	0.07** (0.01)	0.06* (0.08)	0.06* (0.08)	0.06* (0.08)	0.06* (0.08)
Aggregate ESG	0.52 (0.18)		1.13 (0.15)		-0.72 (0.13)		-1.53 (0.11)	
Community Relations		0.33* (0.08)		0.66* (0.06)		0.21 (0.23)		0.28 (0.30)
Corporate Governance		-0.06 (0.40)		-0.26 (0.30)		0.06 (0.39)		0.09 (0.43)
Diversity		-0.14 (0.25)		-0.44 (0.14)		0.16 (0.24)		0.15 (0.35)
Employee Relations		0.27* (0.09)		0.64** (0.05)		-0.26* (0.07)		-0.53* (0.07)
Environment		-0.14 (0.30)		-0.29 (0.31)		-0.36 (0.13)		-0.68 (0.14)
Human Rights		-0.11 (0.38)		-0.25 (0.34)		-0.59* (0.10)		-1.07* (0.10)
Product Safety		-0.03 (0.44)		-0.08 (0.42)		-0.32* (0.09)		-0.59* (0.10)
Intercept	0.37 (0.40)	0.63 (0.34)	-0.13 (0.47)	0.59 (0.37)	3.97** (0.02)	4.78*** (0.00)	4.59** (0.01)	6.10*** (0.00)
Mean Adj. R^2 Cross-section	0.13	0.14	0.13	0.14	0.13	0.13	0.13	0.13
Max. Adj. R^2 Cross-section	0.43	0.43	0.43	0.43	0.37	0.38	0.38	0.38
N	639	639	639	639	639	639	639	639
T	132	132	132	132	60	60	60	60

P-values in parentheses. *$p <= 0.10$, **$p <= 0.05$, ***$p <= 0.01$.

N=average number of cross-sectional observations. T=number of monthly observations.

Mean (Max) Adjusted R^2 is the average (maximum) adjusted R^2s of the T cross-sectional estimations.

TABLE 5. Fama-MacBeth estimation of Model (2) over July 1992-June 2003 and July 2003-June 2008, with industry dummies; the ESG variables are computed using Relative and KO methods. The dependent variable, excess returns, is in percentage points (%).

for product safety in 2003-2008. The aggregate ESG variable, while changing sign between the two periods was not statistically significant in either of them. More interestingly, the employee relations effect changed sign (and statistically significant) from positive during 1992-2003 to negative during 2003-2008. The economic magnitude of its effect is also substantial, shifting from 0.58 - 0.69% per annum in the earlier period to -0.56% or -0.64% per annum later (also for a one-standard-deviation increase in the score).

Next, Equation (3) was Garch-in-Mean estimated for the factor-mimicking portfolios for each of the ESG scores found to have an impact on stock returns, i.e., community relations and employee relations for 1992-2003, as well as human rights, product safety, and employee relations for 2003-2008 (Table 6). For simplicity, the Relative and Relative Best-in-Class aggregation methods for the ESG scores are used due to the strong similarity noted between theirs and their KO counterparts' behavior. The Best-in-Class method yields scores for which it is feasible to build the mutually-exclusive mimicking portfolios at 30th- and 70th-percentile cutoff points. However, with the Relative scores the 20th and 80th cutoff points had to be used for community relations and employee relations scores; 5th and 95th percentiles for human rights, and 15th and 85th percentiles for product safety.

The estimation results are presented in Table 6. Of all the candidate risk-factor mimicking portfolios, only employee relations during 2003-2008 met the necessary condition of a positive relation between its conditional mean and conditional variance using either Relative or Best-in-Class scores (Table 6, second column). The positive sign of the relation indicates a positive premium for firms with low employee scores, i.e., a non-sustainability premium with respect to employee relations risk-factor, which confirms the second working hypothesis and the risk-factor scenario. Low employee relations scores are associated with high non-sustainability risk and therefore high expected returns. However, the intercept in the conditional mean equation is statistically different from zero, which suggests there are some idiosyncratic also driving the returns of the employee-relations portfolio.

A similar message emerges from Garch-in-Mean estimation on the employee-relations portfolio orthogonal to the Market, Small-Minus-Big, and High-Minus-Low portfolios (Table 7). Orthogonalization is performed just as a robustness check of previous findings as it helps correct for correlation between the candidate risk-factor and other risk factors (see Charoenrook and Conrad, 2005). The p-value of δ estimate and its sign indicate a strong positive association between its conditional mean and conditional variance in each of the specifications. However the intercept in the conditional mean equation still carries some extra information besides the conditional variance, which is weaker when using the Best-in-Class method. This could be an indication that some of the information which the intercept carries could be related to industry differences.

While there are reasons to further investigate the possibility that the negative effect of employee relations on returns during the later period might be due to compensation for non-sustainability risk, its positive effect in the earlier period was clearly not compensation for risk. For the earlier period, the necessary condition is not met, as the p-value of the δ coefficient was higher than conventional levels (Table 6, first column). At the same time, the intercept (mu) estimate was negative, though not statistically significant, which is however in line with the implication of the cross-sectional approach of higher unconditional mean return for high employee performers than for the low employee performers. Thus the employee relations effect changed from positive, potentially consistent with mispricing, to negative, possibly representing compensation for risk.

		1992-2003				2003-2008				1992-2008			
		Relative		Relative BC		Relative		Relative BC		Relative		Relative BC	
		p-val	sign	p-val	sign	p-val	sign	p-val	sign	p-val	sign	p-val	sign
Community	delta	0.38	+	0.35	+	0.33	+	0.47	-	0.32	+	0.40	+
Relations	mu	0.17	-	0.28	-	0.43	-	0.48	+	0.22	-	0.21	-
Employee	delta	0.22	+	0.32	+	0.00***	+	0.05**	+	0.29	+	0.28	+
Relations	mu	0.17	-	0.27	-	0.00***	-	0.09*	-	0.27	-	0.34	-
Human	delta	0.34	+	0.45	+	0.15	-	0.26	-	0.15	+	0.38	+
Rights	mu	0.24	-	0.11	-	0.11	+	0.40	+	0.17	-	0.10*	-
Product	delta	0.39	+	0.30	+	0.13	-	0.18	-	0.48	+	0.31	+
Safety	mu	0.44	-	0.42	+	0.14	+	0.13	+	0.48	-	0.37	+
T		132		132		60		60		192		192	

$*p <= 0.10$, $**p <= 0.05$, $***p <= 0.01$. T=number of monthly observations.

TABLE 6. P-value and sign for the conditional variance (delta) and intercept (mu) parameter estimates in the conditional mean equation of the Garch-in-Mean model (Equation 3) for risk-factor mimicking portfolios over July 1992-June 2003, July 2003-June 2008 and July 1992-June 2008. The risk-factor mimicking portfolios are built as Low minus High ESG scores, computed using both Relative and Relative Best-in-Class methods. In the Relative Best-in-Class approach, 30th- and 70th-percentile breakpoints were used; in the Relative method, 20th- and 80th-percentiles for community relations and employee relations; 5th- and 95th-percentiles for human rights; and 15th- and 85th-percentiles for product safety.

Please note that the sample size for the two subperiods is not too big, of respectively 132 and 60 observations, which might increase estimation error, especially for the Garch-in-Mean estimates. Therefore, the results for the subperiods should be interpreted with caution.

In order for employee relations to have behaved as a risk factor in the later period, the sufficient condition of a Sharpe ratio estimate of 0.17 per month must also be met. Our Sharpe ratio estimates for this factor vary depending on the method used to compute the score, but are all substantially lower than the anticipated value, i.e., six or even ten times lower than the benchmark (Table 7, bottom line), meaning a 0.026 (Relative BC) or 0.017 (Relative) risk premium per month. In conclusion, as only the necessary condition for employee relations being a risk factor was met, there is some evidence that the higher returns earned by the low-employee-relations firms was compensation for risk.

The other factors, i.e., community relations in 1992-2003, and human rights and product safety in 2003-2008, did not satisfy the necessary condition to be risk-factors, i.e., the p-value of their δ estimates was higher than a conventional 0.10 (Table 6, first and second columns). On the other hand, the unconditional mean estimate (mu) was positive for both human-rights and product-safety mimicking portfolios and negative for the community relations on the respective periods, though not statistically significant at conventional levels either. This, however, is in line with the cross-sectional estimates of lower returns for the high human-rights and product-safety performers and higher returns for the high community-performers, which might be due to mispricing.

Thus, the benefits of good community relations might have outweighted their costs, but the costs of good human rights and product safety might have outweighted their benefits.

Finally, the corporate governance, diversity, and environment scores had no statistically significant effects on risk-adjusted returns, neither during the overall period nor during the two subperiods. This seems to indicate either that information on these dimensions was efficiently reflected in stock returns or that it was not relevant at all, which is in line with the no effect scenario.

Thus, we have found only weak evidence that employee relations could function as a risk factor (in the second period), and some evidence that the few other indicators' effects on stock returns were due to mispricing.

	Relative				Relative BC			
		Orthogonal to				Orthogonal to		
	Plain	Market	SMB	HML	Plain	Market	SMB	HML
Mean Equation								
intercept	-1.31**	-1.66*	-1.50**	-1.18**	-2.58*	-2.71**	-2.13*	-2.70*
	(0.00)	(0.10)	(0.00)	(0.02)	(0.10)	(0.03)	(0.10)	(0.07)
delta	1.52***	1.79*	1.67***	1.16**	2.29*	2.24***	2.02**	2.33**
	(0.00)	(0.05)	(0.00)	(0.01)	(0.05)	(0.01)	(0.04)	(0.03)
Conditional Variance Equation								
intercept	0.35***	0.33**	0.43***	0.00	0.18	0.00	0.13	0.13
	(0.00)	(0.04)	(0.00)	(0.50)	(0.16)	(0.50)	(0.39)	(0.26)
ARCH1	0.49***	0.63***	0.44***	0.90***	0.83***	0.98***	0.86**	0.87***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.02)	(0.00)
GARCH1	0.25***	0.09*	0.24***	0.14***	0.03*	0.02*	0.04	0.03*
	(0.00)	(0.09)	(0.00)	(0.00)	(0.10)	(0.08)	(0.13)	(0.07)
T	60	60	60	60	60	60	60	60
Sharpe Ratio (%)	1.66	1.95	1.85	1.30	2.59	2.64	2.23	2.67

P-values in parentheses. *$p <= 0.10$, **$p <= 0.05$, ***$p <= 0.01$. T=number of monthly observations.

TABLE 7. Parameter estimates in the conditional mean and conditional variance equation of the Garch-in-Mean model in Equation (3) for the employee-relations risk-factor mimicking portfolio, both plain and orthogonal to the Market, Small-Minus-Big size (SMB) and High-Minus-Low (HML) risk-factors, over July 2003-June 2008. The monthly returns of the employee relations portfolio are in percentage points. The employee relations score was computed using the Relative and Relative Best-in-Class (Relative BC) methods. In the Relative Best-in-Class method, 30th- and 70th-percentile breakpoints were used; in the Relative method, 20th and 80th percentiles.

6. Summary and Conclusions

During the period July 1992-June 2008, only community relations had a positive effect on stock returns, potentially due to mispricing. The benefits of having good community relations might have outweighted their costs, but this might not have been properly incorporated into stock prices.

Some evidence of mispricing is also provided for the positive estimated effect of employee relations on stock returns during July 1992-June 2003, and the weak negative estimated effects of human rights and product safety on stock returns during July 2003-June 2008. The negative estimated effects for human rights and product safety indicate that their benefits might actually have been lower than their costs, but again this was not properly incorporated into stock prices.

There is weak evidence, however, that the change in the effect of employee relations from positive during June 1992-July 2003 to negative during July 2003-June 2008 could be due to compensation for risk. That is, firms with low scores in employee relations had higher expected stock returns than firms with high employee relations scores. This can be explained that firms with low scores in employee relations carry either a non-sustainability risk premium or a ”neglect” premium (by the discriminatory-tastes argument). This change in effect could be a consequence of better public information with respect to employee relations. However, during July 1992-June 2003, the market might not have fully incorporated the net benefits of high employee relations.

Several problems with similar studies have been avoided here. First, in order to capture the actual relationship between ESG and stock returns, which might otherwise be dictated by the unobserved industry classification that could drive both ESG and stock returns, industry-specific effects were controlled for (and they were found to matter). The primary interest was thus in how ESG relates to stock returns within industries, not in the spurious relationship of returns and ESG across industries.

Second, improving over two frequently used aggregation methods, a new aggregation method for measuring ESG was carefully designed and applied.

Third, because many empirical studies show that not all ESG dimensions are equally relevant for stock returns, or, more importantly, that there may be confounding effects among them, the effects of seven ESG dimensions – community relations, corporate governance, diversity, employee relations, environment, human rights, and product safety – were analyzed separately.

These findings may be important for both investors and corporate strategists. Investors may be interested in new evidence that ESG performance is value relevant or that certain non-sustainability risks might exist, while firms might find that they can reduce their cost of capital by promoting certain ESG concerns.

APPENDIX A. "GROUPING TECHNIQUE"

A problem when using the Fama-MacBeth two-pass procedure is estimation bias due to measurement error (sampling variance) in $\widehat{\beta}_j$, the errors-in-variables (EIV) problem (Black et al., 1972). The estimated $\widehat{\beta}_j$ in Equation 1 equals the true (unobservable) beta β_j plus a measurement error (or sample variance) v_j, assumed to be i.i.d. with zero mean and constant variance σ_v^2.

Therefore the residuals θ_j in Equation 1 also contain v_{jt}. OLS estimation leads to an inconsistent estimate of $\widehat{\gamma}_1$ due to the correlation between $\widehat{\beta}_j$ and the residuals θ_j. Even for large samples, as long as the variance of the measurement error σ_v^2 is positive, the estimated coefficient $\widehat{\gamma}_1$ will not converge to the true parameter γ_1 and will be biased towards zero, so-called attenuation bias (Greene, 2003, :85).

A solution developed by Black et al. (1972) and used by Fama and French (1992), the "grouping technique", provides N-consistent estimates of $\widehat{\gamma}_1$. It aims to reduce the variance in the residuals ϵ_j such that, when the sample size goes to infinity, this variance goes to zero.

Securities are first ordered in portfolios based on their individual beta estimates from the time series regressions

$$R_{jt} = \alpha_j + \beta_j R_{mt} + \epsilon_{jt} \tag{7}$$

where R_{jt} is the excess return on asset j, month t; R_{mt} is the excess return on a market index, month t; $\alpha_j = E(R_j) - \beta_j E(R_m)$; $\beta_j = cov(R_{jt}, R_{mt})/var(R_{mt})$; and ϵ_j is independent of R_m and has a normal distribution with zero mean and constant variance.[7]

Portfolio excess return series are then constructed by taking the average of the excess returns in each portfolio. The estimation of Equation (7) - but now at portfolio level - yields *portfolio betas* instead of individual betas. In this case, as the individual error terms ϵ_j are by standard assumption i.i.d. in time and across assets, their averaging within portfolios results in a variance that is proportionally reduced by the number of securities in each portfolio. Thus, when the number of securities goes to infinity and the number of portfolios is kept fixed, the variance of the error terms converges to zero. The estimated coefficient $\widehat{\gamma}_1$ will then be N-consistent. Fama and MacBeth (1973), Fama and French (1992) and Amihud et al. (1992) acknowledge the benefits of this technique in providing more precise β_j estimates and used it in their empirical tests.

Due to their high correlation with the true but unobservable betas, individual beta estimates are used as ranking criteria. Firm-betas have been shown to be highly correlated with size as well (Fama and French, 1992), and therefore both size and individual beta estimates were used as ranking criteria.

For each year, all stocks were ranked in portfolios by firm size and individual β estimates. The individual β values were then estimated with the time series regression Equation (7) on 2 to 5 years of monthly observations prior to June of year t. By ranking the securities in 10 portfolios

[7]Because the market return includes asset i, the disturbances cannot be independent of R_m. However, Jensen (1969) shows that the resulting bias is extremely small (Friend and Blume, 1970).

based on market value as of December 31 of year $t - 1$, and then separating each size-decile into five individual β breakpoints, 50 equally-weighted portfolios were formed as of June 30 of year t, and their returns were then computed for the next 12 months, through June of year $t+1$. Portfolio βs were estimated for each year based on the full remaining period and then assigned to the firms in that portfolio. Fama and French (1992) argue that "the precision of the full-period post-ranking portfolio βs, relative to the imprecise β estimates that would be obtained for individual stocks, more than makes up for the fact that the true βs are not the same for all stocks in a portfolio." At the same time, they note that variation in a firm's β is ensured by the fact that, due to changes in size or individual β estimate, a firm can change portfolios from year to year.

Appendix B. KLD's ESG Strength and Weakness Indicators

COMMUNITY STRENGTHS. Charitable Giving. The company has consistently given over 1.5% of trailing three-year net earnings before taxes (NEBT) to charity, or has otherwise been notably generous in its giving. Innovative Giving. The company has a notably innovative giving program that supports nonprofit organizations, particularly those promoting self-sufficiency among the economically disadvantaged. Companies that permit nontraditional federated charitable giving drives in the workplace are often noted in this section as well. Non-US Charitable Giving. The company has made a substantial effort to make charitable contributions abroad, as well as in the U.S. To qualify, a company must make at least 20% of its giving, or have taken notably innovative initiatives in its giving program, outside the U.S. 4. Support for Housing. The company is a prominent participant in public/private partnerships that support housing initiatives for the economically disadvantaged, e.g., the National Equity Fund or the Enterprise Foundation. Support for Education. The company has either been notably innovative in its support for primary or secondary school education, particularly for those programs that benefit the economically disadvantaged, or the company has prominently supported job-training programs for youth. Other Strength. The company has either an exceptionally strong volunteer program, in-kind giving program, or engages in other notably positive community activities. WEAKNESSES. Investment Controversies. The company is a financial institution whose lending or investment practices have led to controversies, particularly ones related to the Community Reinvestment Act. Negative Economic Impact. The companys actions have resulted in major controversies concerning its economic impact on the community. These controversies can include issues related to environmental contamination, water rights disputes, plant closings, put-or-pay contracts with trash incinerators, or other company actions that adversely affect the quality of life, tax base, or property values in the community. Other Weakness. The company is involved with a controversy that has mobilized community opposition, or is engaged in other noteworthy community controversies.

CORPORATE GOVERNANCE STRENGTHS. Limited Compensation. The company has recently awarded notably low levels of compensation to its top management or its board members. The limit for a rating is total compensation of less than \$500,000 per year for a CEO or \$30, 000 per year for outside directors. Ownership Strength. The company owns between 20% and 50% of another company KLD has cited as having an area of social strength, or is more than 20% owned by a firm that KLD has rated as having social strengths. When a company owns more than 50% of another firm, it has a controlling interest, and KLD treats the second firm as if it is a division of the first. Other Strength. The company has an innovative compensation plan for its board or executives, a unique and positive corporate culture, or some other initiative not covered by other KLD ratings. WEAKNESSES. High Compensation. The company has recently awarded notably high levels of compensation to its top management or its board members. The limit for a rating is total compensation of more than \$10 million per year for a CEO or \$100, 000 per year for outside directors. Tax Disputes. The company has recently been involved in major tax disputes involving more than \$100 million with the Federal, state, or local authorities. Ownership Weakness. The company owns between 20% and 50% of a company KLD has cited as having an area of social weakness, or is more than 20% owned by a firm KLD has rated as having areas of weakness. When a company owns more than 50% of another firm, it has a controlling interest, and KLD treats the second firm as if it is a division of the first. Other Weakness. The company restated its earnings over an accounting controversy, has other accounting problems, or is involved with some other controversy not covered by other KLD ratings.

DIVERSITY STRENGTHS. CEO. The companys chief executive officer is a woman or a member of a minority group. Promotion. The company has made notable progress in the promotion of women and minorities, particularly to line positions with profit-and-loss responsibilities in the corporation. Board of Directors. Women, minorities, and/or the disabled hold four seats or more (with no double counting) on the board of directors, or one-third or more of the board seats if the board numbers less than 12. Work/Life Benefits. The company has outstanding employee benefits or other programs addressing work/life concerns, e.g., childcare, elder care, or flextime. Women & Minority Contracting. The company does at least 5% of its subcontracting, or otherwise has a demonstrably strong record on purchasing or contracting, with women- and/or minority-owned businesses. Employment of the Disabled. The company has implemented innovative hiring programs, other innovative human resource programs for the disabled, or otherwise has a superior reputation as an employer of the disabled. Gay & Lesbian Policies. The company has implemented notably progressive policies toward its gay and lesbian employees. In particular, it provides benefits to the domestic partners of its employees. Other Strength. The company has made a notable commitment to diversity that is not covered by other KLD ratings. WEAKNESSES. Controversies. The company has either paid substantial fines or civil penalties as a result of affirmative action controversies, or has otherwise been involved in major controversies related to affirmative action issues. Non-Representation. The company has no women on its board of directors or among its senior line managers. Other Weakness. The company is involved in diversity controversies not covered by other KLD ratings.

EMPLOYEE RELATIONS STRENGTHS. Cash Profit Sharing. The company has a cash profit sharing program through which it has recently made distributions to a majority of its workforce. Employee Involvement. The company strongly encourages worker involvement and/or ownership through stock options available to a majority of its employees, gain sharing, stock ownership, sharing of financial information, or participation in management decision-making. Health and Safety Strength. The company is noted by the US Occupational Health and Safety Administration for its safety programs. Retirement Benefits Strength. The company has a notably strong retirement benefits program. Union Relations. The company has a history of notably strong union relations. Other Strength. The company has strong employee relations initiatives not covered by other KLD ratings. WEAKNESSES. Union Relations. The company has a history of notably poor union relations. Health and Safety Weakness. The company recently has either paid substantial fines or civil penalties for willful violations of employee health and safety standards, or has been otherwise involved in major health and safety controversies. Workforce Reductions. The company has reduced its workforce by 15% in the most recent year or by 25% during the past two years, or it has announced plans for such reductions. Retirement Benefits Weakness. The company has either a substantially underfunded defined benefit pension plan, or an inadequate retirement benefits program. Other Weakness. The company is involved in an employee relations controversy that is not covered by other KLD ratings.

ENVIRONMENT STRENGTHS. Beneficial Products and Services. The company derives substantial revenues from innovative remediation products, environmental services, or products that promote the efficient use of energy [costa], or it has developed innovative products with environmental benefits. (The term environmental service does not include services with questionable environmental effects, such as landfills, incinerators, waste-to-energy plants, and deep injection wells.) Clean Energy. The company has taken significant measures to reduce its impact on climate change and air pollution through use of renewable energy and clean fuels or through energy efficiency. The company has demonstrated a commitment to promoting climate-friendly policies and practices outside its own operations. Communications. The company is a signatory to the CERES Principles, publishes a notably substantive environmental report, or has notably effective internal communications systems in place for environmental best practices. Pollution Prevention. The company has notably strong pollution prevention programs including both emissions reductions and toxic-use reduction programs. Recycling. The company either is a substantial user of recycled materials as raw materials in its manufacturing processes, or a major factor in the recycling industry. Other Strength. The company has demonstrated a superior commitment to management systems, voluntary programs, or other environmentally proactive activities. WEAKNESSES. Hazardous Waste. The companys liabilities for hazardous waste sites exceed $50 million, or the company has recently paid substantial fines or civil penalties for waste management violations. Regulatory Problems. The company has recently paid substantial fines or civil penalties for violations of air, water, or other environmental regulations, or it has a pattern of regulatory controversies under the Clean Air Act, Clean Water Act or other major environmental regulations. Ozone Depleting Chemicals. The company is among the top manufacturers of ozone depleting chemicals such as HCFCs, methyl chloroform, methylene chloride, or bromines. Substantial Emissions. The companys legal emissions of toxic chemicals (as defined by and reported to the EPA) from individual plants into the air and water are among the highest of the companies followed by KLD. Agricultural Chemicals. The company is a substantial producer of agricultural chemicals, i.e., pesticides or chemical fertilizers. Climate Change. The company derives substantial revenues from the sale of coal or oil and its derivative fuel products, or the company derives substantial revenues indirectly from the combustion of coal or oil and its derivative fuel products. Such companies include electric utilities, transportation companies with fleets of vehicles, auto and truck manufacturers, and other transportation equipment companies. Other Weakness. The company has been involved in an environmental controversy that is not covered by other KLD ratings.

HUMAN RIGHTS STRENGTHS. Indigenous Peoples Relations Strength. The company has established relations with indigenous peoples near its proposed or current operations (either in or outside the U.S.) that respect the sovereignty, land, culture, human rights, and intellectual property of the indigenous peoples. Labor Rights Strength. The company has outstanding transparency on overseas sourcing disclosure and monitoring, or has particularly good union relations outside the U.S. Other Strength. The company has undertaken exceptional human rights initiatives, including outstanding transparency or disclosure on human rights issues, or has otherwise shown industry leadership on human rights issues not covered by other KLD human rights ratings. WEAKNESSES. Burma Weakness. The company has operations or investment in, or sourcing from, Burma. Labor Rights Concern. The companys operations outside the U.S. have had major recent controversies related to employee relations and labor standards or its U.S. operations have had major recent controversies involving sweatshop conditions or child labor. Indigenous Peoples Relations Weakness. The company has been involved in serious controversies with indigenous peoples (either in or outside the U.S.) that indicate the company has not respected the sovereignty, land, culture, human rights, and intellectual property of indigenous peoples. Other Weakness. The companys operations outside the U.S. have been the subject of major recent human rights controversies not covered by other KLD ratings.

PRODUCT STRENGTHS. Quality. The company has a long-term, well-developed, company-wide quality program, or it has a quality program recognized as exceptional in U.S. industry. R&D/Innovation. The company is a leader in its industry for research and development (R&D), particularly by bringing notably innovative products to market. Benefits to Economically Disadvantaged. The company has as part of its basic mission the provision of products or services for the economically disadvantaged. Other Strength. The companys products have notable social benefits that are highly unusual or unique for its industry. WEAKNESSES. Product Safety. The company has recently paid substantial fines or civil penalties, or is involved in major recent controversies or regulatory actions, relating to the safety of its products and services. Marketing/ Contracting Controversy. The company has recently been involved in major marketing or contracting controversies, or has paid substantial fines or civil penalties relating to advertising practices, consumer fraud, or government contracting. Antitrust. The company has recently paid substantial fines or civil penalties for antitrust violations such as price fixing, collusion, or predatory pricing, or is involved in recent major controversies or regulatory actions relating to antitrust allegations. Other Weakness. The company has major controversies with its franchises, is an electric utility with nuclear safety problems, defective product issues, or is involved in other product related controversies not covered by other KLD ratings.

REFERENCES

Ahn DH, Conrad J, Dittmar RF. 2005. Basis assets. *Review of Financial Studies* **22**: 5133–5174.

Amihud Y, Christensen BJ, Mendelson H. 1992. Further evidence on the risk-return relationship. *Research Paper, Stanford University* **1248**: [Accessed 19.04.2010].

Arbel A, Carvell S, Strebel P. 1983. Giraffes, institutions and neglected firms. *Financial Analysts Journal* **39**: 2–8.

Barnea A, Rubin A. 2006. Corporate social responsibility as a conflict between shareholders. *EFA 2006 Zurich Meetings* **Available at** $http://papers.ssrn.com/sol3/papers.cfm?abstract\_id = 686606$: [Accessed 19.04.2010].

Bauer R, Koedijk K, Otten R. 2005. International evidence on ethical mutual fund performance and investment style. *Journal of Banking and Finance* **29**: 1751–1767.

Becchetti L, Ciciretti R. 2006. Corporate social responsibility and stock market performance. *CEIS Working Paper* **79**: [Accessed 19.04.2010].

Black F, Jensen M, Scholes M. 1972. The capital asset pricing model: Some empirical tests. In Jensen M (ed.) *Studies in the Theory of Capital Markets*. Praeger Publishers, 72–194.

Charoenrook A, Conrad J. 2005. Identifying risk-based factors. *AFA 2006 Boston Meetings Paper* **Available at** $http://papers.ssrn.com/sol3/papers.cfm?abstract\_id = 757210$: [Accessed 19.04.2010].

Chatterji A, Brooks C, Pavelin S. 2008. How well do social ratings actually measure corporate social responsibility? *Journal of Economics and Management Strategy* **18**: 125–169.

Clark G, Knight E. 2009. Institutional investors, the political economy of corporate disclosure, and the market for corporate environmental and social responsibility: Implications from the uk companies act. *U. of Pennsylvania Journal of Business Law* **11**: 259–296.

Cochrane J. 2005. *Asset Pricing*. Princeton University Press.

Derwall J, Gunster N, Bauer R, Koedijk K. 2005. The ecoefficiency premium puzzle. *Financial Analysts Journal* **61**: 61–63.

Derwall J, Verwijmeren P. 2007. Csr and the cost of equity capital. In Derwall J (ed.) *The Economic Virtues of SRI and CSR*. Haveka, 193–213.

Dhaliwal D, Li OZ, Tsang A, Yang YG. 2009. Voluntary non-financial disclosure and the cost of equity capital: the case of corporate social responsibility reporting. *Working Paper* **Available at:** $http://papers.ssrn.com/sol3/papers.cfm?abstract\_id = 1343453$: [Accessed 19.04.2010].

Dowell G, Hart S, Yeung B. 2000. Do corporate global environmental standards create or destroy market value? *Management Science* **46**: 1059–1074.

Dufresne F, Savaria P. 2004. Corporate social responsibility and financial risk. *Journal of Investing* **13**: 57–66.

Edmans A. 2008. Does the stock market fully value intangibles? employee satisfaction and equity prices. *Working paper Series at the R.L. White Center for Financial Research* **16-07**:

[Accessed 19.04.2010].

Epstein M, Freedman M. 1994. Social disclosure and the individual investor. *Accounting, Auditing and Accountability Journal* **7**: 94–109.

Fama E, French K. 1992. The cross-section of expected returns. *Journal of Finance* **47**: 426–465.

Fama E, French K. 1993. Common risk factors in the returns of stocks and bonds. *Journal of Financial Economics* **33**: 3–56.

Fama E, MacBeth J. 1973. Risk, return and equilibrium: Empirical tests. *Journal of Political Economy* **81**: 607–636.

Friend I, Blume M. 1970. Measurement of portfolio performance under uncertainty. *American Economic Review* **60**: 561–575.

Galema R, Plantinga A, Scholens B. 2008. The stocks at stake: return and risk in socially responsible investment. *Journal of Banking and Finance* **32**: 2646–2654.

Greene WH. 2003. *Econometric Analysis*. NJ Prentice Hall.

Haigh M, Hazelton J. 2004. Financial markets: A tool for social responsibility? *Journal of Business Ethics* **52**: 59–71.

Heinkel, Kraus A, Zechner J. 2001. The effect of green investment on corporate behavior. *Journal of Financial and Qantitative Analysis* **36**: 431–449.

Hong H, Kacperczyk M. 2009. The price of sin: the effects of social norms on markets. *Journal of Financial Economics* **93**: 15–36.

Jegadeesh N, Titman S. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *The Journal of Finance* **48**: 65–91.

Jensen M. 1969. Risk, the pricing of capital assets, and the evaluation of investment portfolios. *Journal of Business* **42**: 167–247.

Kempf A, Osthoff P. 2007. The effect of socially responsible investing on portfolio performance. *European Financial Management* **13**: 908–922.

Kinder P. 2005. Socially responsible investing: an evolving concept in a changing world **Available at** $http://www.kld.com/resources/papers/SRIevolving050901.pdf$: [Accessed 19.04.2010].

KPMG. 2008. International survey of corporate responsibility reporting **Available at** $http://us.kpmg.com/RutUS\_prod/Documents/8/Corporate\_Sustainability\_Report\_US\_Final.pdf$: [Accessed 19.04.2010].

MacKinley A. 1995. Multifactor models do not explain deviations from the capm. *Journal of Financial Economics* **38**: 3–28.

Merton R. 1987. A simple model of capital market equilibrium with incomplete information. *Journal of Finance* **42**: 483–510.

Patten D. 1990. The market reaction to social responsibility disclosures: The case of the sullivan principles signing. *Accounting Organisations and Society* **15**: 515–87.

Renneboog L, Horst JT, Zhang C. 2008. Socially responsible investments: institutional aspects, performance, and investor behavior. *Journal of Banking and Finance* **32**: 1723–1742.

SIF. 2007. 2007 report on socially responsible investing trends in the united states, social investment forum - excutive summary **Available at** $http://www.socialinvest.org/pdf/SRI\_Trends\_ExecSummary\_2007.pdf$: [Accessed 19.04.2010].

Statman M. 2006. Socially responsible indexes. *Journal of Portfolio Management* **32**: 100–109.

Statman M, Glushkov D. 2009. The wages of social responsibility. *Financial Analysts Journal* **65**: 33–46.

Wall L. 1995. Some lessons from basic finance for effective socially responsible investing. *Economic Review - Federal Reserve Bank of Atlanta* **80**: 1–12.

Paper IV

Doing worse while doing less good [a]

Cristiana Manescu[b]

Abstract

This paper investigates how annual abnormal returns react to current and past rating revisions in corporate responsible behavior in a panel data spanning 16 years. I find that increases in less responsible behavior led to persistent negative abnormal returns, which were particularly strong for the area of corporate governance, and weaker for product safety and the environment. These results are robust to concerns of endogeneity, i.e., that the negative stock price movements would lead to an update in the areas of social responsibility concerns. In contrast, increases in already strong responsible behavior did not generate a systematic reaction in stock returns.

JEL: C23, C26, G14, M14

Keywords: corporate responsible behavior, stock returns, control functions approach

---

[a] *This research was conducted as part of the research programme Sustainable Investments, funded by the Swedish Foundation for Strategic Environmental Research, MISTRA.*

[b] Center for Finance, School of Business, Economics and Law, University of Gothenburg, Box 600, Gothenburg, SE-40 530. Email: cristiana.manescu@cff.gu.se. I am grateful to my supervisors, Martin Holmen and Catalin Starica, and also to Mans Soderbom and Hossein Asgharian (discussant at Final Seminar) for valued input; as well as to seminar participants at 3rd PRI conference Copenhagen 2010, and at the Mistra workshop Rome 2010. The usual disclaimer applies.

1. Introduction

The concept of responsible and sustainable investing has attracted a lot of attention from both the investor community and the academic community. While responsible investment might mean different things to different people, it generally refers to incorporating into an investment strategy some from a number of environmental, social, and governance (ESG) issues.

In an attempt to answer one of this decade's controversial questions of whether responsible investing is profitable or financially harmful, the academic literature has put forward various theoretical hypotheses as well as empirical findings based on a multitude of empirical designs and ESG measurement data. The consensus seems to be that responsible investing is most of the time profitable rather than harmful but also that the direction of causation might go both ways.

There are a multitude of studies generally showing a positive relationship between corporate responsible involvement in levels and stock returns, even when controlling for various pricing factors (Kempf and Osthoff, 2007; Statman and Glushkov, 2009). However, this relationship can merely be an association if the issue of endogeneity, i.e., the extent to which past firm performance might also be driving socially responsible behavior, is not properly addressed. The literature using balance-sheet measures of performance has made the first steps toward addressing the endogeneity issue and showed that an apparent positive relation between CSR and profitability is to a large extent due to firm-specific effects (e.g., Elsayed and Paton, 2005). However, causality has been generally elusive in studies analyzing the relationship between stock returns and ESG performance, with the exception of several event studies (e.g., Becchetti et al., 2009; Krueger, 2009). Therefore, this study aims to fill this gap by investigating in greater detail the *mechanism* through which revisions in ratings of firms' social responsibility rating are reflected into their unexpected stock return. I will pursue this aim by (1) using *changes* in ESG performance instead of levels to capture the dynamic effect of ESG performance; (2) using a longitudinal cross-sectional analysis in order to control for various confounding factors while maintaining an event-study flavor; (3) taking advantage of the unique way Kinder, Lydenberg, and Domini (KLD) Research and Analytics rates ESG performance by distinguishing between *areas of strength* and *areas of weakness* within each of the seven ESG dimensions and not aggregate these indicators by dimension; and (4) explicitly correcting for a potential feedback from unexpected returns to ESG performance in the next period.

By analyzing the effect of *revisions in ESG scores*, rather than that of their levels, I can provide an analysis complementary to the prevailing positive documented relation between many of KLD rated dimensions and risk-adjusted stock returns. If high absolute KLD scores lead to high stock returns, then it is to be expected that an increase in the KLD score should lead to an increase in stock returns as well; if not, the relationship found between KLD scores and stock returns could be due to underlying characteristics of the firm that overlap the KLD score and which have not been properly taken into account in the analysis. This motivation is similar to one of the explanations proposed by Gompers et al. (2003) for the positive risk-adjusted returns reaction to strong shareholder rights. One of their hypotheses is that "governance provisions did not cause poor performance but rather were correlated with other characteristics that were associated with abnormal returns" in the analyzed period. Therefore, by using revisions in ESG performance rather than levels, I can measure ESG performance impact directly, assuming that the other characteristics do not simultaneously change.

Further, evaluating the *separate impact* that revisions in areas of ESG strength and ESG weakness might have on stock returns could reveal whether it is the positive ESG performance that is rewarded in the market or the negative ESG performance that is penalized, an important question that affects the incentives of corporations. Whereas in many similar studies, also based on KLD data, ESG performance is usually aggregated into one number by subtracting the number of weaknesses from the number of strengths, there is evidence pointing out that the market reaction is stronger to weaknesses than it is to strengths (e.g., Galema et al., 2008), which could partly be due to the documented fact that weaknesses provide a more accurate indication of future risks (Chatterji et al., 2008). For example, van der Laan et al. (2008) analyze disaggregated KLD data in relation to return on assets (ROA) and earnings per share (EPS). They find a strong negative effect of weakness indicators but a zero effect of strength indicators, with the only exception of environmental strengths that are found being negatively related to both ROA and EPS. Moreover, using disaggregate ESG data in a *panel data* setting provides the advantage of measuring the effect of each indicator net of the impact of the other indicators. This characteristic is absent in studies based on portfolio methodology (buying high-ESG performers and selling the low-ESG performers), which fail to account for correlations between the ESG dimensions, i.e., firms that are good in one dimension could also be good in a different dimension. At the same time, while an event study methodology can typically also capture direct effects of revisions in ESG performance, it fails to control for other factors that might affect returns, which is possible in a panel-data setting.

Finally, and potentially most importantly, the reverse causality issue is analyzed even further by explicitly trying to correct for the effect that abnormal past financial performance (past abnormal returns) might have on revisions in ESG performance. Studies documenting higher financial performance for firms with higher ESG engagement are often criticized for not properly accounting for the endogeneity between high financial performance and ESG performance.

Many studies using KLD ratings of U.S. firms show that portfolios built on various KLD dimensions (as well as an aggregate KLD measure) earn abnormal returns over long periods of time even when accounting for the Carhart four risk factors (e.g., Kempf and Osthoff, 2007; Statman and Glushkov, 2009, etc.). However, when the analysis is performed at the firm level, by analyzing the direct impact of KLD scores on either the ex-ante cost of equity capital or the ex-post stock returns, mixed results are found with respect to certain ESG issues. Whereas Derwall and Verwijmeren (2007) find a positive impact of a social index and a negative impact of product, corporate governance and environment indicators over 1992-2006, Manescu (2010) identifies a positive impact of community and employee relations indicator over June 1992-July 2003 and a negative impact of employee relations, environment and human rights over July 2003-June 2008, leading to the implication that portfolio analyses and cross-sectional analyses might reveal different aspects of the relationship between ESG performance and stock performance.

The empirical analyses based on other data providers provide similar findings. A portfolio of US stocks with high rating on an aggregate environment indicator evaluated by Innovest Strategic Value Advisors is found to earn substantially higher average returns than its low counterpart during the period from 1995-2003 (Derwall et al., 2005). Using UK microdata on environmental, employment and community activities in 2002 provided by the Ethical Investment Research Service (EIRIS), Brammer et al. (2006) find that environmental and employment indicators are negatively correlated to returns, whereas the community indicator is weakly positively related. von Arx and Ziegler (2008) analyze the reaction of US and European stock returns to environmental performance and a social index reflecting relationships with several stakeholders as provided by Sarasin&Cie in 2002. Both dimensions are found to have a positive impact on stock returns but a weaker one in the European sample.

The results of this study show that only ESG weaknesses determine a reaction in abnormal returns, one that is negative, whereas revisions to strength indicators generate zero effect. However, despite an estimated overall negative effect of weakness revisions of 4.61%, it is only those with respect to governance- particularly strong and those with respect to product quality and environment that also have an individual causal effect. Using an aggregate measure of weakness revision would have led to the conclusion that all weaknesses are relevant, whereas diversity, employee relations and human rights are in fact not.

The remaining of the paper is organized as follows: Section 2 presents the ESG and financial data, Section 3 briefly presents the model and Section 4 presents baseline estimation results. In Section 5, the endogeneity treatment is explained, whose results are discussed in Section 6. Finally, Section 7 briefly discusses several robustness checks and Section 8 concludes.

2. Data

ESG performance is here measured by the Kinder, Lydenberg and Domini (KLD) Research and Analytics ratings KLD STATS, which is a dataset with snap-shots of performance along seven dimensions: community relations, diversity, corporate governance, employee relations, environment, human rights and product safety. Each ESG area is graded *annually* based on a number of positive and negative indicators, i.e., strengths and weaknesses, which are given a score of 1 if present and 0 otherwise. For a list of these indicators, please see Manescu (2010); for their in-depth description, see Becchetti and Ciciretti (2009).

KLD data are available for an average of 650 companies every year during 1991-2000; since 2003, these data have been available for over 3,000 companies. Despite data availability, to limit the impact of a sudden increase in sample size *during* the analysis period as well as to minimize the risk of 0 rating meaning no rating, I will mainly focus on the DS400 and SP400 components. Statman and Glushkov (2009) adopt a similar strategy for the same reason. Moreover, summary statistics (unreported) indicate that when moving from the larger to the smaller sample, the number of firms with annual revisions in strengths or weaknesses diminishes only marginally compared to the total number of firms which almost halves, which justifies the use of the smaller sample.

For the purpose of this paper, there will be fourteen ESG variables for each strength and each weakness indicator in the seven dimensions. Revisions in ESG performance are measured as a discrete variable, entailing the yearly change in the number of strengths and weaknesses. Summary statistics in Table 1 indicate that the percentage of firms recording a change in any KLD dimension, strengths or weaknesses, is usually low, ranging between 0.5% (human rights strengths) and 21% (diversity strengths).[1] Another observation that will prove helpful for the later analysis is that changes in areas of weakness occur roughly 30% more often than changes in areas of strength.

Please recall that ESG strengths and weaknesses were not aggregated on purpose along the seven dimensions because of the suspicion that they may be structurally different with opposed impacts on returns which aggregation might obscure. For example, areas of strength like *charitable giving* and areas of weakness like *anti-trust violations* – both cost increasing – might have a similar (negative) impact on returns. To gain support for this claim, I performed a factor analysis on the 14 variables (in levels) and estimated whether variation in these indicators could be broadly summarized by a handful of common underlying factors. The findings indicate that at least seven factors were required to explain a significant portion of total variation (of less than 50%). Also, there was no clear pattern as to how these factors load on the various strengths and weaknesses except that they seem to be generally loading uniquely on individual indicators. This is evidence that the

[1] The disproportionately low percentage of firms recording a change with respect to human rights strengths is driven to a large extent by the fact that most of human rights strengths were introduced only after 2000.

Variable	Obs.	Mean	Std. Dev.	Min	Max	
Abnormal return (%)	10066	0.80	36.82	-145.25	196.54	
Size (ln(Assets))	10066	15.41	1.71	9.73	21.51	
Leverage (LTD/T_Capital)	10066	32.87	22.92	0	149.81	
Financials	10066	0.13	0.34	0	1	
Basic Materials	10066	0.08	0.27	0	1	
Industrials	10066	0.19	0.39	0	1	
Technology	10066	0.09	0.29	0	1	
Consum.Goods	10066	0.14	0.35	0	1	
Oil&Gas	10066	0.05	0.22	0	1	
Healthcare	10066	0.07	0.26	0	1	
Consum. Serv	10066	0.15	0.36	0	1	
Telecom	10066	0.02	0.15	0	1	
Utilities	10066	0.07	0.26	0	1	
Variable	Obs.	Mean	Std. Dev.	Min	Max	% change in total
deltacomstr	9082	0.01	0.36	-3	3	11
deltacomwkn	9082	0.02	0.24	-2	2	6
deltacgovstr	9082	0.01	0.26	-2	2	6
deltacgovwkn	9082	0.05	0.46	-2	3	20
deltadivstr	9082	0.09	0.53	-3	4	21
deltadivwkn	9082	0.01	0.31	-2	2	9
deltaempstr	9082	0.03	0.39	-2	2	13
deltaempwkn	9082	0.05	0.48	-3	3	20
deltaenvstr	9082	0.02	0.3	-2	4	7
deltaenvwkn	9082	0.02	0.37	-3	3	11
deltahumstr	9082	0	0.07	-1	1	1
deltahumwkn	9082	0	0.22	-2	2	5
deltaprostr	9082	0	0.19	-1	2	4
deltaprowkn	9082	0.03	0.35	-4	2	10
deltasumstr	9082	0.16	0.93	-4	7	63
deltasumwkn	9082	0.19	0.96	-7	7	81

Due to rounding error, the sum over subindicators may not add up exactly.

TABLE 1. Summary Statistics for firms in SP500 or DS400 during 1991-2007. "% change in total" indicates the percentage of firms that recorded a change in the respective dimension.

indicators are rather heterogeneous, which therefore supports the choice of maintaining the disaggregate dimensions throughout the analysis.

The abnormal return for each security (i), each year(t), ($AbnRet_{i,t}$) is determined as the difference between its actual return ($ActRet_{i,t}$) and its expected return ($ExpRet_{i,t}$). The more comprehensive four-factor model was used to calculate expected returns. Technically, the security expected return for year t is the risk-free rate ($RiskFreeR_t$) plus the product of the security sensitivities to the four risk factors, i.e. the market portfolio and the size, value and momentum risk-factor-mimicking portfolios, and the expected risk premiums on these factors in year t (see Equation 2). The expected risk premiums on these factors are their realized returns in year t. The actual return is the buy-and-hold return for year t (i.e., price on December 31st in year t over price on December 31st in year $t-1$). Buy-and-hold returns have been recommended in studies where returns are measured over a longer-horizon (several months) because additive cumulation procedures (such as cumulative abnormal return) are systematically positively biased due to the bid-ask spread (e.g., Conrad and Kaul, 1993). However, this comes at the cost of a more highly right-skewed distribution than that of cumulative returns (Kothari and Warner, 1997).

$$\begin{aligned} AbnRet_{i,t} &= ActRet_{i,t} - ExpRet_{i,t} \qquad (1)\\ ExpRet_{i,t} &= RiskFreeR_t + \hat{\beta}^{Market}_{i,t-1} ExpRet_{(Market_t - Rf_t)} + \hat{\beta}^{Size}_{i,t-1} ExpRet_{(Small_t - Big_t)} \\ &+ \hat{\beta}^{Value}_{i,t-1} ExpRet_{(High_t - Low_t)} + \hat{\beta}^{Mom}_{i,t-1} ExpRet_{(Winners_t - Loosers_t)}, \qquad (2) \end{aligned}$$

where $t = 1991...2007$ and $i = 1...N$, N is the number of securities in each year.

The sensitivities $\hat{\beta}^k_{i,t-1}, k = 1...4$ are the multiple regression coefficients of returns on the four factors, Model 3, estimated on a sample of 60 (or at least 24) monthly returns ending in December year $t-1$. Therefore, each $\hat{\beta}^k$ captures the security's sensitivity to risk factor k net the impact of the others.

$$\begin{aligned} ActRet_{i,s} - Rf_s &= \alpha_i + \beta^{Market}_{i,t-1} ActRet_{(Market_s - Rf_s)} + \beta^{Size}_{i,t-1} ActRet_{(Small_s - Big_s)} \\ &+ \beta^{Value}_{i,t-1} ActRet_{(High_s - Low_s)} + \beta^{Mom}_{i,t-1} ActRet_{(Winners_s - Loosers_s)} + \gamma_{i,s}, \qquad (3) \end{aligned}$$

where $s = 1...60$ months up to December year $t-1$.

Moreover, the $\beta^k_{i,t-1}, k = 1...4$, parameters are estimated by Model 3 such that they satisfy the constraint that the expected return on security i cannot be lower than -100%, meaning that you cannot expect to lose more than your entire investment. Specifically, when estimating regression 3, the constraint on $\beta^k_{i,t-1}, k = 1...4$, parameters is that their product with the expected risk premiums on the risk factors next period t, i.e., the expected return component in excess of the risk-free rate, cannot go below -100%.

The summary statistics of the beta estimates in Table 2 indicate that both Market and Size betas are tightly distributed around their mean, whereas the other two are more dispersed. Also, the mean estimate market beta of 1 is in line with expectations.

Finally, if Model 3 is well specified, the risk-factor sensitivities $\hat{\beta}^k_i, k = 1...4$ should meet the following condition: if, by assumption, one owns the entire stock market, then one

	(1) Mean	(2) SD	(3) p25	(4) p75	(5) $\Sigma_{j=1}^{N}\hat{\beta}_j * w_j$
Market beta	1.01	0.03	0.99	1.03	1.01
Size beta	0.65	0.07	0.60	0.69	-0.05
Value beta	0.21	0.13	0.12	0.35	0.04
Momentum beta	-0.13	0.05	-0.18	-0.09	-0.05

TABLE 2. Mean, standard deviation (SD), 25th (p25) and 75th(p75) percentiles of the yearly average beta sensitivities to the for risk factors (Columns 1 to 4). The yearly average of the market capitalization weighted sum of the four securities betas (Column 5).

should earn no more and no less than the return on the market portfolio. That is equivalent to taking the sum over all securities on the left-hand side of equation 3, weighted by their market capitalization ratio, and obtaining the market portfolio excess return on the right hand side of this equation. More specifically, the following conditions should hold every year: $\Sigma_{j=1}^{N}\hat{\beta}_j^{Market} * w_j = 1$, $\Sigma_{j=1}^{N}\hat{\beta}_j^{Size} * w_j = 0$, $\Sigma_{j=1}^{N}\hat{\beta}_j^{Value} * w_j = 0$, $\Sigma_{j=1}^{N}\hat{\beta}_j^{Momentum} * w_j = 0$, where w_j is the market weight of security j. Column 5 in Table 2 indicates that these conditions hold on average and therefore equation 2 is well specified and can be used to estimate expected excess returns.

The other control variables are:

- Year dummies
- Industry dummies (nine) based on an MSCI industry classification of 10 (oil & gas, basic resources, industrials, consumer services, healthcare, consumer goods, telecommunications, utilities, financials and technology), whose distribution is presented in Table 1,

and, at a later stage, a set of instruments consisting of firm size (measured by natural logarithm of assets) and firm leverage (measured as the ratio of long-term liabilities to total capital; common equity was used instead of total capital as standardization variable, but firms often had very small value for equity leading to unusually high ratios and therefore it was dropped.)

After the panel data construction for period 1991-2007, several outlier elimination was performed sequentially: abnormal returns are required to be larger than -150% and lower than 200% (0.006% observations lost), leverage is imposed to be positive and lower than 150% (another 0.005% observations lost)

In Table 1, summary statistics are presented for variables of interest in the sample consisting of SP500 and DS400 components.

3. Empirical Strategy

The ESG data, begins in 1991 and spans sixteen years. My main use of this data here is to test whether revisions in ESG performance determine an *unexpected* adjustment in stock prices. There is already a body of evidence showing that ESG ranking and stock returns tend to be positively correlated (Kempf and Osthoff, 2007), but this association has not yet been proven to be causal except in the limited setting of event studies (e.g., Krueger, 2009). The unexpected return is the difference between actual return and the four-factor expected return, defined by Equations 1 and 2.

The explanatory variables of interest are the changes in ESG performance, *separated* by each dimension and by areas of strength and weakness, which are measured both contemporaneously and with a lag. There are two channels through which corporate ESG engagement can impact stock returns, and therefore timing is important. One is through how investors/research analysts react to news (unexpected information) on the ESG performance (information availability factor); the other is through its (unexpected) impact on underlying firm cash flows (economic factor).

By the information availability factor, the KLD scores pertaining to year t are made public late January, early February year $t+1$. However, it is only the compiled data that are made available at this time point, as throughout the year, KLD research analysts irregularly disseminate newsletters containing references to publicized incidents of social responsibility *as they occur*, which largely contributes to the revisions compiled at the end of the year. [2] This justifies estimating the contemporaneous relation between overall revisions to the KLD scores and unexpected returns over the same period, which would largely capture a reaction to ESG news estimate. The yearly unexpected return is meant to capture the cumulative effect of the various KLD revisions over the period, despite the fact that it almost certainly also reflects impacts of other events. However, this would affect our results only by making them weaker- if we manage to capture some effect, then the actual effect would be even stronger. Still, in the context of the information availability factor, a lagged relationship between ESG revisions and unexpected returns can capture any return reaction to the dissemination of compiled data, which, in this form, might reveal overall firm information unnoticed by the newsletters.

By the economic factor, the economic benefits of ESG performance could materialize within the same period or one period later, even if the costs are almost entirely borne upfront. The change in ESG performance can be seen as a permanent shock to cash-flows that might impact returns during the same period or with a lag. For example, increases in productivity as a result of investing in employee training could result in unexpected returns a period later. This also motivates the inclusion of the lagged changes in ESG performance among the explanatory variables. There might be some effects of ESG engagement in periods later than second one, but they should be fairly small.

[2] Between August 2001 and April 2007, KLD newsletters were sent out on average every three weeks (Krueger, 2009).

The seven ESG dimensions are deliberately not aggregated in order to control for possible conflicting effects. Therefore, the baseline model is as follows:

$$AR_{i,t} = a + \Sigma_{j=1}^{7} b_j \Delta Str_{i,t}^{j} + \Sigma_{j=1}^{7} c_j \Delta Wkn_{i,t}^{j} + \Sigma_{j=1}^{7} d_j \Delta Str_{i,t-1}^{j} + \Sigma_{j=1}^{7} e_j \Delta Wkn_{i,t-1}^{j} + \\ + \Sigma_{k=1}^{9} f_k Industry_k + \Sigma_{k=1}^{15} g_k Year_k + \epsilon_{i,t} \quad (4)$$

where $AR_{i,t}$ is the yearly abnormal return, $\Sigma_{j=1}^{7} \Delta Str_{i,t}^{j}$ and $\Sigma_{j=1}^{7} \Delta Wkn_{i,t}^{j}$ are, respectively, yearly changes in the number of strengths and weaknesses in each of the seven KLD dimensions.

In theory, no other control variable would be justified in the model for abnormal returns (Model 4) because in a four-factor world, as assumed when computing the abnormal returns, it is these four factors alone that can impact returns, and any other factor should have zero systematic impact. These potentially additional factors are captured by the $\epsilon_{i,t}$, which is normally distributed with mean zero and a diagonal covariance matrix. However, there is evidence (e.g., Fama and French, 1997) that these four factors still do not capture all of the between-industry or time variation in returns. For this reason, industry and year dummies were added to the baseline model.

Model 4 will be estimated with Fixed Effects or Random Effects - the Hausman test will indicate which one is preferable. Alternatively, pooled Ordinary Least Squares (OLS) with clustered errors can also be used.

Economically and statistically significant coefficients on changes in strengths and weaknesses at time t and $t-1$ would reconfirm the value relevance of ESG performance. In line with prevailing empirical evidence that high KLD scores lead to high returns, I would expect a positive significant coefficient on some of the strengths variables and a negative coefficient on some of the weaknesses variables.

Therefore, the main tests are as follows:

(1) the joint net effect of changes in strengths indicators, current and lagged, is positive
(2) the joint net effect of changes weaknesses indicators, current and lagged, is negative

Which strength and weakness variables might turn out to have a significant impact on returns depends on how sensitive investors are to the respective information they carry as well as on which theoretical framework of ESG engagement prevails.

It is sometimes claimed that the relationship between financial and ESG performance can in fact go the other way , i.e., from financial performance to ESG performance. In such a case, the ESG variables in Model 4 would be endogenous, thus leading to biased OLS estimates. I allow for endogeneity in Section 5 using a control function approach.

4. Estimates of the impact of changes in the KLD scores on abnormal returns

In this section, I report various estimations of the impact of changes in ESG performance on abnormal returns, ignoring the reverse causality issue. I will allow for endogenous ESG performance in Section 5. I estimate Model 4 with pooled OLS (and clustered standard errors at the firm level), fixed effects (FE) and random effects (RE). Three main patterns emerge.

First, the pooled OLS, RE and FE deliver similar results of Model 4's estimation, and the Hausman test indicates that the RE estimation is consistent ($p-value = 0.36$). Therefore, I use the RE estimates as main estimates. This verified consistency of the RE estimates can also be interpreted as a confirmation of the strict exogeneity of firm specific effects, i.e., they are uncorrelated with the error term. This means that I might have succeeded in not having any omitted variables in the error term of the model for abnormal returns, Model 4; or, that the four-factor model used to estimate abnormal returns succeeds in capturing all relevant information for stock returns (with the caveat of correcting for between industry variation in a second step).

Second, as anticipated, revisions in strength and weakness areas trigger different adjustment reaction in returns. Therefore, no revision in ESG strength has any significant impact on returns, either during the same period or with a lag. Taken jointly, changes in strengths in all dimensions have *zero net effect* on returns - according to the Wald test (Panel B, Table 3), which is consistent across estimations. This has very strong implications, meaning that improvement or worsening in already positive ESG performance does not generate any reaction in abnormal returns. The good news is that getting extra strength does not harm returns, and the bad news is that it does not lead to positive abnormal returns either, as claimed by many proponents of ESG. KLD strength indicators were not found to have an impact on returns either in revision form in the event study analysis of Krueger (2009) or in level form (for the employee relations indicator only)in the cross-sectional analysis of Galema et al. (2008).

On the other hand, revisions to weakness dimensions *do have a joint impact* on abnormal returns, both with immediate and delayed effect (Panel B, Table 3). The overall return adjustment to changes in all weaknesses is *negative*, consistent across estimations (the null hypothesis of negative summated effect of changes in weaknesses at time $t-1$ and t is comfortably accepted). Therefore, worsening one's already poor/negative ESG standing is penalized by the market. Acquiring an extra weakness (in any dimension) reduces the abnormal return by an estimated, significant 0.66% (RE) per year. Similarly, acquiring an extra weakness in all seven dimensions, either during the last or current period, reduces the abnormal return by 4.62% (RE model). However, this is very rare, as there is only one case in the 16 years of data. This magnitude is comparable, to abnormal return premiums

	(1) (P-OLS)	(2) (FE)	(3) (RE)
Panel A			
Δ_t Community Strengths	-0.29	-1.07	-0.56
	(0.76)	(0.33)	(0.60)
Δ_t Governance Strengths	-1.88	-2.37	-2.19
	(0.23)	(0.13)	(0.14)
Δ_t Diversity Strengths	0.17	-0.64	-0.15
	(0.81)	(0.38)	(0.83)
Δ_t Employee Strengths	0.99	0.89	0.86
	(0.37)	(0.39)	(0.38)
Δ_t Environment Strengths	-0.87	-1.29	-1.10
	(0.36)	(0.33)	(0.38)
Δ_t Human Strengths	-1.59	-2.17	-2.02
	(0.77)	(0.69)	(0.69)
Δ_t Product Strengths	-1.76	-1.58	-1.65
	(0.41)	(0.47)	(0.43)
Δ_t Community Weaknesses	**-2.29***	-2.15	-2.33
	(0.09)	(0.18)	(0.13)
Δ_t Governance Weaknesses	**-2.17****	**-2.26****	**-2.25*****
	(0.01)	(0.01)	(0.01)
Δ_t Diversity Weaknesses	0.18	0.35	0.24
	(0.89)	(0.79)	(0.84)
Δ_t Employee Weaknesses	-0.31	-0.18	-0.24
	(0.70)	(0.83)	(0.76)
Δ_t Environment Weaknesses	-0.61	-1.07	-0.81
	(0.51)	(0.32)	(0.43)
Δ_t Human Weaknesses	0.87	0.21	0.60
	(0.57)	(0.90)	(0.71)
Δ_t Product Weaknesses	**-1.69***	**-1.89***	**-1.76***
	(0.06)	(0.09)	(0.10)
Δ_{t-1} Community Strengths	0.33	-0.42	0.08
	(0.75)	(0.70)	(0.94)
Δ_{t-1} Governance Strengths	-1.55	-2.12	-1.86
	(0.29)	(0.18)	(0.22)
Δ_{t-1} Diversity Strengths	-0.13	-0.86	-0.43
	(0.86)	(0.26)	(0.55)
Δ_{t-1} Employee Strengths	1.45	1.09	1.32
	(0.20)	(0.30)	(0.19)
Δ_{t-1} Environment Strengths	-1.05	-1.70	-1.30
	(0.37)	(0.20)	(0.30)
Δ_{t-1} Human Strengths	3.45	3.10	3.24
	(0.41)	(0.57)	(0.54)
Δ_{t-1} Product Strengths	-1.88	-1.35	-1.76
	(0.40)	(0.52)	(0.38)

	(1) (P-OLS)	(2) (FE)	(3) (RE)
Δ_{t-1} Community Weaknesses	-1.56	-1.43	-1.61
	(0.29)	(0.39)	(0.31)
Δ_{t-1} Governance Weaknesses	1.07	0.71	0.96
	(0.21)	(0.43)	(0.27)
Δ_{t-1} Diversity Weaknesses	-0.35	-0.45	-0.37
	(0.78)	(0.73)	(0.76)
Δ_{t-1} Employee Weaknesses	0.70	0.86	0.82
	(0.42)	(0.32)	(0.33)
Δ_{t-1} Environment Weaknesses	**-1.60***	**-1.77***	**-1.70***
	(0.07)	(0.10)	(0.10)
Δ_{t-1} Human Weaknesses	-0.05	-0.99	-0.43
	(0.97)	(0.57)	(0.80)
Δ_{t-1} Product Weaknesses	-0.20	-0.47	-0.34
	(0.83)	(0.68)	(0.75)
Year Dummies	yes	yes	yes
Industry Dummies	yes	no	yes
Panel B - Tests (p-values)			
Net zero effect all Strengths(Δ_t and Δ_{t+1})	0.54	0.45	0.50
Net zero effect all Weaknesses ($\Delta_t + \Delta_{t+1}$)	0.08	0.04	0.06
Net negative effect all Weaknesses ($\Delta_t + \Delta_{t+1}$)	0.96	0.98	-
Hausman test RE estimates consistent: 0.36			
Observations	8131	8131	8131
Groups		996	996
R^2	0.05	0.04	.
F-stat (Wald chi2 for RE)	9.16	7.54	360.26
	(0.00)	(0.00)	(0.00)
Panel C - Two aggregated strength and weakness indicators			
Δ_t Sum Strengths	-0.18	**-0.72***	-0.43
	(0.65)	(0.10)	(0.29)
Δ_t Sum Weaknesses	**-1.00*****	**-1.11*****	**-1.06*****
	(0.00)	(0.01)	(0.01)
Δ_{t-1} Sum Strengths	-0.06	-0.61	-0.30
	(0.88)	(0.16)	(0.47)
Δ_{t-1} Sum Weaknesses	0.00	-0.16	-0.05
	(0.98)	(0.71)	(0.90)

Robust p values in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 3. In Panel A, estimation results of Model 4 with Pooled OLS- clustered standard errors (Column 1), panel fixed effects (2) and random effects (3). In panel B, p-values of several post-estimation Wald tests and of the Hausman test on the consistency of the Random effects estimates. In Panel C, estimation results for an identical model except that strength and weakness indicators are each aggregated. The dependent variable, abnormal stock returns is measured in percentage points (%).

of e.g., 4.90% or 5.21% found by Kempf and Osthoff (2007) during 1992-2004, or 5.54% found by Statman and Glushkov (2009) during 1992-2007, both papers using a best-in-class strategy and a four-factor model, despite other methodological differences. Moreover, as strength revisions were found not to have any impact on returns, this findings might indicate that their results were primarily driven by the impact of weaknesses.

My evidence on the impact of revisions in both strengths and weakness indicators on unexpected returns confirms earlier findings based on different methodologies (Galema et al., 2008; Krueger, 2009). One explanation for this observed effect can be provided in the lines of the "content analysis" findings by Krueger (2009). Using specific linguistics techniques, he finds that negative events' descriptions (related to weaknesses), provided in the KLD newsletter, contain more quantitative information than positive ones (related to strengths). Therefore, because weakness indicators measure significantly more tangible and measurable quantities than strength indicators do, investors might find it easier to assess their impact on share value. Strength indicators, on the other hand, seem to reflect qualities that might have only a long-term effect on share value (e.g., the company has strong health and safety programs or the company has notably strong pollution prevention programs), even though the empirical analysis did not reveal any second period effects of strength indicators on returns.

Finally, one can observe strong return adjustments to several individual weakness dimensions in particular. An increase in the number of governance weaknesses has an immediate negative effect on abnormal returns. Here, the governance weaknesses category reflects the following: high compensation (more than $10 million a year for a CEO or $100,000 a year for outside directors), whether the company owns between 20% and 50% of a company KLD has cited as having an area of social weakness or is more than 20% owned by a firm KLD has rated as having areas of weakness. Since 2006, it has also reflected accounting controversies, weak reporting on a wide range of social and environmental performance measures and involvement with noteworthy controversies on public policy issues, etc.). Given the already established relevance of corporate governance practices for stock returns (Gompers et al., 2003), it seems that investors and research analysts have real time access to information of the kind captured by our governance weakness measure.

Other weakness indicators that were found to affect abnormal returns during the same period are those with respect to product safety, whereas those with respect to the environment materialize during the next period (see Table 3). However, their effect is weaker than that of governance. It can be noticed from the KLD descriptions that weaknesses more often reflect areas of economic loss or financial penalty that is specifically related to ESG, whereas strengths may represent more elusive, non-quantified qualities. Therefore, it may not come as a surprise that the reaction in unexpected returns follows only with respect to weaknesses.

This analysis finds no reaction in unexpected returns to revisions in areas of weaknesses with respect to diversity, employee relations, or human rights. Whereas under the caveat expressed in Section 3, this could partly be due to the length of the period which unexpected returns are measured; it might also indicate that none of these weaknesses carry any relevant information for share value. In fact, Krueger (2009) does not find any systematic return reaction to either news on diversity or human rights but does find a negative reaction to negative news related to employee relations.

As weakness and strength subcomponent indicators tend to have a consistent impact on unexpected return, I have also estimated Model 4 with two aggregated weakness and strength indicators instead of the fourteen individual ones. The results presented in Panel C in Table 3 confirm that only contemporaneous weaknesses have a significant impact on returns. It is now obvious that by aggregation one would have missed the weak effect of environment and might have implied an equal weighting of all subcompoenents when the individual results indicate that only few subcomponents have an impact on returns.

5. Endogenous changes in ESG performance

In this section, I worry that the direction of causation may not be from changes in ESG performance to unexpected returns, as modeled in Equation 4, but rather may be the other way, which would bias all our estimates. This could happen for two reasons. First, unexpected good or bad price movements may draw the attention of KLD analysts to revise the rating of the respective firm. Second, firms' prior financial performance might determine its future engagement with ESG issues. It has been claimed, for example, that financially successful firms are more likely to improve their ESG standing, as they have the required resources for it. Therefore, I report additional results that should be more robust than the OLS estimates to these potential problems.

The traditional way to correct for endogeneity is by using adequate instruments that are correlated with the endogenous variable (i.e., the fourteen changes in KLD scores) but not correlated with the residual, ϵ_{it}. Lag (past) abnormal return is the first candidate, given the initial concern that it might affect current changes in ESG performance. Firm size and firm risk (leverage) would also make good instruments, as they have been shown to be correlated with ESG performance (Waddock and Graves, 1997). Industry dummies usually explain variation in ESG performance and therefore should be included among instruments; however, they are not excluded instrument, being also part of the abnormal return equation.

In addition to these shared instruments, to ensure identification, each endogenous variable needs to be instrumented with at least one unique instrument (there must be at least as many excluded instrument as endogenous variables). A common practice in panel data analyses is to use earlier lags of the endogenous variable as their own instruments. In this case, I can use the number of strengths or weaknesses in levels at time $t-3$, $Str_{k,i,t-3}$ and $Wkn_{k,i,t-3}$, to instrument, respectively, the change in the number of strengths or weaknesses at time t, i.e., $\Delta Str_{k,i,t}$ and $\Delta Wkn_{k,i,t}$.[3] Finally, as I used the extended four-factor model to account for any systematic variation in the expected component of returns, none of these instruments should be correlated with the residual ϵ_{it} in the unexpected return equation 4.

Two stage least squares (2SLS) is the typical estimation approach used to address endogeneity. With 2SLS, I would have to first estimate fourteen first-stage regressions, modeling each component of ESG strengths and weaknesses separately, and then use the predictions instead of the actual values in the second stage. However, given the large number of endogenous variables and given that the ESG strengths and weaknesses are not fundamentally different, nor are the instruments I found, at least not within the two groups, identification would be difficult to achieve in practice. Ideally, it would be preferable to use a more flexible approach that would allow reducing the number of first-stage regressions. Recall that there are fourteen endogenous variables in the model precisely because I want to explicitly model the separate impact on stock returns of revisions in every ESG strength and weakness dimension. Therefore, I want to have the fourteen separate variables in the main equations, but I would also like to meaningfully reduce the number of first-stage regressions. One such flexible approach is the two-stage control function approach (Söderbom et al., 2006)[4].

With the control function approach, without any loss of generality, I can reduce the number of first-stage regressions from fourteen to two by aggregating all strengths and all weaknesses into two distinct indicators, which would then be regressed on the corresponding instruments in the first stage regressions:

$$\begin{aligned} SumChangeStr = m_1 + n_1 AR_{i,t-1} + \Sigma_{k=1}^{7} p_1^k Str_{i,t-3}^k + r_1 Size_{i,t-1} + \\ + s_1 Leverage_{i,t-1} + \Sigma_{k=1}^{9} t_k^1 Industry_k + u_{i,t}^{str} \end{aligned} \tag{5}$$

$$\begin{aligned} SumChangeWkn = m_2 + n_2 AR_{i,t-1} + \Sigma_{k=1}^{7} p_2^k Wkn_{i,t-3}^k + r_2 Size_{i,t-1} + \\ + s_2 Leverage_{i,t-1} + \Sigma_{k=1}^{9} t_k^2 Industry_k + u_{i,t}^{wkn}, \end{aligned} \tag{6}$$

where $SumChangeStr = \Sigma_{k=1}^{7} \Delta Str_{i,t}^k$ is the sum of changes in the seven strengths, while $SumChangeWkn = \Sigma_{k=1}^{7} \Delta Wkn_{i,t}^k$ is the sum of changes in the seven weaknesses.

In the second-stage regression, I would include the estimated residuals $\hat{u}_{i,t}^{str}$ and $\hat{u}_{i,t}^{wkn}$, in current and lagged forms, respectively, among the explanatory variables of the main Model 4. The abnormal return abnormal returns equation would be re-written as

[3]Even though the second lag in levels might seem as the first choice for an instrument, due to its implicit presence in the main equation through lagged changes in the ESG variables, and thus implicit suspicion of endogeneity, I will use the third lag in levels as the first suitable lag for instrumentation.
[4]For an in-depth analysis of the control function approach please see Card (2001)

$$AR_{i,t} = a + \Sigma_{j=1}^{7} b_j \Delta Str_{j,i,t} + \Sigma_{j=1}^{7} c_j \Delta Wkn_{j,i,t} + \Sigma_{j=1}^{7} d_j \Delta Str_{j,i,t-1} + \Sigma_{j=1}^{7} e_j \Delta Wkn_{j,i,t-1} + \\ + \Sigma_{k=1}^{9} f_k Industry_k + \Sigma_{k=1}^{15} g_k Year_k + \alpha_1 \hat{u}_{i,t}^{str} + \beta_1 \hat{u}_{i,t}^{wkn} + \alpha_2 \hat{u}_{i,t-1}^{str} + \beta_2 \hat{u}_{i,t-1}^{wkn} + \epsilon_{i,t}. \tag{7}$$

It can be shown starting from Equation 1 that the OLS estimates of $\Sigma_{j=1}^{7} b_j$ and $\Sigma_{j=1}^{7} c_j$ in Equation 7 are identical to the 2SLS estimates (see, Wooldridge, 2007, p. 1-4).

A simple test for the null of exogeneity is a Wald test (heteroskedasticity-robust) if the coefficients of the first-stage residuals terms in Equation 7, $\alpha_1, \alpha_2, \beta_1$ and β_2, are jointly zero.[5]

6. Results

The main concern in assuming the changes in the KLD scores to be endogenous is that they might be correlated with past abnormal returns, in which case the OLS coefficient estimates in Table 3 could be biased. Simple estimated correlations between changes in scores and abnormal returns at time $t-1$ or time $t-2$ indicate that past abnormal returns are rather *uncorrelated* with current changes in KLD scores, with few exceptions (Table 4). Therefore, the endogeneity concern may not be a real problem in this setup. In any case, I proceed with a formal test of endogeneity.

Correlation coefficients between levels at time $t-3$ and changes at time t for the 14 ESG strength and weakness indicators are in the range -.03, -.16 and statistically significant (Table 4), with only one zero correlation coefficient. This supports the use of own lags in levels as instruments. Moreover, these negative correlations for all indicators suggest that firms with a high number of either strengths or weaknesses tend to migrate toward lower scores, i.e. bad firms tend to get better and good firms tend to get worse. This is in line with dynamics of KLD scores identified by Goss and Roberts (2009). Otherwise, the dynamics between the various ESG strengths and weaknesses do not seem to follow any obvious pattern. One can note, however, that areas of weakness tend to be positively associated with positive changes in areas of strength three periods later, more than twice as often as the frequency at which areas of strength are associated with positive changes in areas of weakness. This might indicate that firms also tend to compensate for certain weaknesses also by acquiring strengths in other dimensions.

I continue our endogeneity correction approach by checking whether the excluded instruments, size, leverage, past abnormal returns and two-period-lagged strengths and weaknesses, respectively, have any explanatory power for the endogenous variables, a necessary condition for identification. Based on the first-stage regressions, i.e., Regressions 6 and 7, I test for the joint significance of the coefficients on the excluded instruments (mentioned

[5]This corresponds to the Hausman endogeneity test in the 2 Stage Least Squares framework.

above). For both equations, I can safely reject the hypothesis that these coefficients are jointly zero ($p - value_{EXCRES} = 0.00$ in both cases, Table 5).

Having established this, I next test for whether the coefficients on the residuals' terms from the first-step equations are jointly zero in the main equation 7. The Wald test safely accepts the null hypotheses of exogeneity ($p - value_{EXOGEN} = 0.48$, Table 5). Therefore, I have provided additional evidence that the negative impact on returns of changes in ESG weaknesses is causal and cannot be attributed to endogeneity concerns.

	Δcomstr	Δcomwkn	Δcgovstr	Δcgovwkn	Δdivstr	Δdivwkn	Δempstr
ar_{t+1}	0.00	-0.01	-0.02	0.02	0.00	0.00	0.02
ar	-0.01	-0.02	-0.01	-0.03*	0.00	0.01	0.01
L.ar	0.00	0.01	-0.02*	0.01	0.01	0.01	0.01
L3.comstr	**-0.10***	0.02	0.05*	0.01	0.05*	0.02*	0.01
L3.comwkn	0.02	**-0.07***	0.05*	0.01	0.00	-0.01	0.02*
L3.cgovstr	-0.01	0.00	**-0.09***	0.00	-0.01	-0.01	-0.00
L3.cgovwkn	0.03*	0.012	**0.06***	-0.03*	0.05*	0.02*	-0.01
L3.divstr	0.00	0.01	**0.07***	0.02*	-0.02	0.03*	0.02
L3.divwkn	0.01	0.03*	0.01	0.01	-0.01	**-0.15***	-0.01
L3.empstr	0.02	0.01	0.04*	0.00	0.01	-0.00	**-0.13***
L3.empwkn	0.01	0.01	0.03*	0.02	-0.00	-0.01	0.04*
L3.envstr	-0.00	0.02	0.03*	-0.00	-0.02	-0.01	-0.00
L3.envwkn	-0.00	0.04*	**0.06***	0.01	0.00	0.01	0.04*
L3.humstr	-0.03*	0.00	0.01	-0.00	-0.00	-0.00	-0.00
L3.humwkn	0.01	0.00	0.04*	0.01	0.01	0.02	0.04*
L3.prostr	0.02	-0.00	0.03*	-0.00	0.00	-0.00	-0.00
L3.prowkn	0.02*	0.04*	**0.08***	0.03*	0.04*	0.03*	0.02

	Δempwkn	Δenvstr	Δenvwkn	Δhumstr	Δhumwkn	Δprostr	Δprowkn
ar_{t+1}	0.01	-0.02	-0.01	0.01	0.01	-0.01	0.00
ar	0.00	-0.01	0.00	0.00	0.01	0.01	-0.01
L.ar	-0.01	-0.01	0.01	-0.01	0.02*	0.01	0.00
L3.comstr	-0.00	0.01	-0.01	0.01	0.00	0.00	0.03*
L3.comwkn	0.02	0.05*	0.00	0.00	0.01	0.01	0.01
L3.cgovstr	0.01	-0.01	0.02	0.03*	0.01	-0.03*	-0.02
L3.cgovwkn	0.01	**0.06***	-0.01	0.02	0.03*	0.01	0.01
L3.divstr	0.02	**0.08***	-0.01	0.01	0.00	-0.01	0.03*
L3.divwkn	0.01	0.01	0.01	0.00	0.01	-0.02	0.00
L3.empstr	-0.00	0.04*	0.00	-0.00	0.02	-0.03*	0.00
L3.empwkn	**-0.11***	0.05*	0.02	0.00	0.01	0.01	-0.01
L3.envstr	0.01	**-0.08***	-0.00	-0.00	-0.01	0.00	-0.02
L3.envwkn	0.00	0.04*	-0.03*	0.00	0.00	0.01	0.01
L3.humstr	-0.00	-0.01	-0.02	-0.05*	0.03*	0.00	0.00
L3.humwkn	-0.00	0.04*	0.02	0.04*	**-0.17***	0.02	-0.00
L3.prostr	-0.01	0.03*	-0.00	0.01	-0.01	**-0.16***	-0.01
L3.prowkn	0.01	**0.06***	0.02	-0.01	-0.01	0.00	**-0.10***

* statistically significant at 5% level. Coefficients in bold are higher than abs(.05).

Table 4. Correlation matrix between selected variables. $ar_t + 1$ refers to abnormal return leading by one period the change in scores. L. and l3. are the lag 1 and lag 3 transforms.

7. Robustness

Several robustness tests have been performed to ensure the validity of our results, which were found among several measurement or methodological alternatives. First, as mentioned, the buy-and-hold procedure for computing abnormal returns reduces the cumulation bias of the cumulative abnormal return procedure but does so at the expense of positive skewness in returns distribution. To investigate the impact that this alternative return calculation might have on the results, I also used it to calculate the dependent variable in Model 4. Histograms of the two approaches to calculate abnormal returns (unreported) confirm that cumulative abnormal returns are almost symmetrically distributed around zero, while the buy-and-hold procedure produces abnormal returns with a distribution skewed to the right.

The estimation results in Table 6 indicate that the results, even though they are qualitatively identical, are statistically stronger when cumulative abnormal returns are used. Acquiring an additional weakness with respect to governance or product safety has a contemporaneous negative effect on returns. Worsened performance with respect to the environment and community leads to a negative return one year later.

Another robustness check that implies using several lags of abnormal returns as instruments for changes in the KLD scores, in addition to one lag, also provides similar results. Finally, the effects are persistent when the enlarged stock universe since 2001 is used in the analysis. All of the robustness results are available upon request.

Tests (p-values)	
EXCRES for $SumChangeStrengths$	19.95
	(0.00)
EXCRES for $SumChangeWeaknesses$	27.35
	(0.00)
EXOGEN ($\alpha_1, \alpha_2, \beta_1$ and β_2)	0.87
	(0.48)

p values in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 5. EXCRES - Wald test for joint significance of the identifying instruments in first-step equations 7 and 6 (the exclusion restrictions); EXOGEN - Wald test for the joint significance of the residual terms in Equation 7.

Panel A	(1) (P-OLS)	(2) (FE)	(3) (RE)
Δ_t Community Concerns	-2.49*	-2.28	-2.49*
	(0.06)	(0.12)	(0.07)
Δ_t Governance Concerns	-2.03***	-2.21***	-2.09***
	(0.01)	(0.01)	(0.01)
Δ_t Diversity Concerns	0.37	0.72	0.47
	(0.75)	(0.53)	(0.67)
Δ_t Employee Concerns	0.00	-0.00	0.03
	(1.00)	(1.00)	(0.97)
Δ_t Environment Concerns	-0.68	-0.97	-0.83
	(0.41)	(0.32)	(0.38)
Δ_t Human Concerns	0.33	-0.27	0.10
	(0.80)	(0.86)	(0.95)
Δ_t Product Concerns	-1.92**	-2.02**	-1.98**
	(0.02)	(0.05)	(0.04)
Δ_{t-1} Community Concerns	-2.52*	-2.39	-2.54*
	(0.07)	(0.11)	(0.08)
Δ_{t-1} Governance Concerns	0.89	0.62	0.82
	(0.24)	(0.45)	(0.29)
Δ_{t-1} Diversity Concerns	-0.29	-0.16	-0.25
	(0.80)	(0.89)	(0.82)
Δ_{t-1} Employee Concerns	0.52	0.65	0.62
	(0.51)	(0.41)	(0.41)
Δ_{t-1} Environment Concerns	-1.73**	-1.80*	-1.80*
	(0.04)	(0.06)	(0.06)
Δ_{t-1} Human Concerns	-0.03	-0.81	-0.32
	(0.98)	(0.61)	(0.83)
Δ_{t-1} Product Concerns	-0.35	-0.52	-0.44
	(0.70)	(0.61)	(0.66)
Year Dummies	yes	yes	yes
Industry Dummies	yes	no	yes
Panel B- Tests (p-values)			
Net zero effect all Strengths(Δ_t and Δ_{t+1})	0.71	0.33	0.60
Net zero effect all Weaknesses ($\Delta_t + \Delta_{t+1}$)	0.02	0.01	0.02
Net negative effect all Weaknesses ($\Delta_t + \Delta_{t+1}$)	0.99	0.99	-
Hausman test RE estimates consistent: 0.43			
R^2	0.05	0.06	.
F-stat (Wald chi2 for RE)	10.04	11.21	522.34
	(0.00)	(0.00)	(0.00)

Robust p values in parentheses. *** p<0.01, ** p<0.05, * p<0.1. N=8131, Groups =996

Table 6. In Panel A, estimation results of Model 4 with Pooled OLS- clustered standard errors (Column 1), panel fixed effects (2) and random effects (3). In panel B, p-values of several post-estimation Wald tests and of the Hausman test on the consistency of the Random effects estimates. The dependent variable, abnormal stock returns is measured in percentage points.

8. Conclusions

This paper explores in greater detail the value relevance of ESG attributes by investigating the abnormal stock return adjustment to revisions in ESG performance in a panel-data analysis with an event study flavor. By working on changes in ESG performance and carefully addressing the reverse-causality issue, this is the first study - to the best of our knowledge - to be taking a closer look at the complex mechanism between ESG performance and stock returns. The data set measuring social responsibility is provided by KLD, which also offers a unique opportunity into assessing separately the impact of areas of positive socially responsible behavior (strengths) and areas of negative socially responsible behavior (weaknesses), as the research firm offers specific indicators for each dimension.

The results are generally weak and do not indicate an abnormal return adjustment from most revisions in ESG performance. In particular, none of the seven strength indicators has any impact on returns, and their estimated overall net impact is zero. Revisions in weaknesses, on the other hand, do have a joint effect on returns that is negative of up to 4.61% for an extra weakness in all seven dimensions. However, this is due to only three individual effects: a contemporaneous effect of revisions in corporate governance and product safety weaknesses and a lagged effect of revisions in environment weaknesses. Moreover, these effects are causal, as a treatment for endogeneity revealed no support for the concern that past performance might impact revisions in ESG performance.

The distinct estimated relationship between revisions in areas of ESG strengths and ESG weaknesses could be due to several reasons (as suggested by Krueger, 2009, as well). First, areas of weaknesses more often reflect a loss expressed in quantitative terms (e.g., the company has recently paid substantial fines for waste management violations, related to the safety of its products, or it has recently awarded notably high levels of compensation to its top management or its board members), whereas areas of strength are usually graded in more qualitative terms and from a long-term perspective (e.g., the company has notably strong pollution prevention programs or the company has outstanding transparency on overseas sourcing disclosure). Therefore, it might be easier for investors to evaluate the impact of ESG weaknesses on share value than it is for them to evaluate the impact of ESG strengths. Second, it might be easier and less costly for KLD to identify negative events than it is for them to identify positive events, given that the negative ones are what typically attracts a high level of media coverage.

The evidence provided for the distinct impacts of ESG weaknesses and strengths indicates that conclusions drawn on an aggregated measure might be misleading due to a lack of homogeneity among components. This study shows that using an aggregate measure of ESG weaknesses revisions could even lead to a generalized conclusion of an overall negative effect of all dimensions on returns, whereas the effect found here was primarily due to three of them.

References

Becchetti, L. and Ciciretti, R. (2009). Corporate social responsibility and stock market performance. *Applied Financial Economics*, 19(16):1283 – 1293.

Becchetti, L., Ciciretti, R., and Hasan, I. (2009). Corporate social responsibility and shareholder's value: an empirical analysis. *Bank of Finland Research Discussion Paper*, 1.

Brammer, S., Brooks, C., and Pavelin, S. (2006). Corporate social performance and stock returns: Uk evidence from disaggregate measures. *Financial Management*, 35(3):75–116.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 69(5):1127–1160.

Chatterji, A., Brooks, C., and Pavelin, S. (2008). How well do social ratings actually measure corporate social responsibility? *Journal of Economics and Management Strategy*, 18(1):125–169.

Conrad, J. and Kaul, G. (1993). Long-term market overreaction or biases in computed returns. *Journal of Finance*, 48:39–63.

Derwall, J., Gunster, N., Bauer, R., and Koedijk, K. (2005). The ecoefficiency premium puzzle. *Financial Analysts Journal*, 61(2):51–63.

Derwall, J. and Verwijmeren, P. (2007). Csr and the cost of equity capital. In Derwall, J., editor, *The Economic Virtues of SRI and CSR*, pages 193–213. Haveka.

Elsayed, K. and Paton, D. (2005). The impact of environmental performance on financial performance: Static and dynamic panel data evidence. *Structural Change and Economic Dynamics*, Available at(16):395–412.

Fama, E. and French, K. (1997). Industry costs of equity. *Journal of Financial Economics*, 43:153–193.

Galema, R., Plantinga, A., and Scholens, B. (2008). The stocks at stake: return and risk in socially responsible investment. *Journal of Banking and Finance*, 32:2646–2654.

Gompers, P., Ishii, J., and Metrick, A. (2003). Corporate governance and equity prices. *The Quarterly Journal of Economics*, pages 107–155.

Goss, A. and Roberts, G. S. (2009). The cost of virtue:corporate social responsibility and the cost of debt financing. *Mimeo*, $http://papers.ssrn.com/sol3/papers.cfm?abstract\_id = 955041$.

Kempf, A. and Osthoff, P. (2007). The effect of socially responsible investing on portfolio performance. *European Financial Management*, 13(5):908–922.

Kothari, S. and Warner, J. B. (1997). Measuring long-horizon security price performance. *Journal of Financial Economics*, 43:301–339.

Krueger, P. (2009). Stakeholder information and shareholder value. *Working Paper*, Available at.

Manescu, C. (2010). Stock returns reaction to environmental social and governance performance: mispricing of compensation for risk? *University of Gothenburg Working Paper*, 372.

Söderbom, M., Teal, F., Wambugu, A., and Kahyarara, G. (2006). The dynamics of returns to education in kenyan and tanzanian manufacturing. *Oxford Bulletin of Economics and Statistics*, 68(3):261–288.

Statman, M. and Glushkov, D. (2009). The wages of social responsibility. *Financial Analysts Journal*, 65(4):33–46.

van der Laan, G., van Ees, H., and van Witteloostuijn, A. (2008). Corporate social and financial performance: an extended stakeholder theory, and empirical test with accounting measures. *Journal of Business Ethics*, 79(3):299–310.

von Arx, U. and Ziegler, A. (2008). The effect of csr on stock performance: New evidence for the usa and europe. *Mimeo*, Available at http://econpapers.repec.org/paper/ethwpswif/08-85.htm.

Waddock, S. and Graves, S. (1997). The corporate social performance- financial performance link. *Strategic Management Journal*, 18(4):303–319.

Wooldridge, J. (2007). What's new in econometrics? lecture 6: Control functions and related methods. *NBER Summer Institute, Available at: http* $: //www.nber.org/WNE/Slides7-31-07/slides\_6\_controlfuncs.pdf$, page [Accessed 20.09.2010].

Previous doctoral theses in the Department of Economics, Gothenburg

Avhandlingar publicerade innan serien Ekonomiska Studier startades
(Theses published before the series Ekonomiska Studier was started):

Östman, Hugo (1911), Norrlands ekonomiska utveckling
Moritz, Marcus (1911), Den svenska tobaksindustrien
Sundbom, I. (1933), Prisbildning och ändamålsenlighet
Gerhard, I. (1948), Problem rörande Sveriges utrikeshandel 1936/38
Hegeland, Hugo (1951), The Quantity Theory of Money
Mattsson, Bengt (1970), Cost-Benefit analys
Rosengren, Björn (1975), Valutareglering och nationell ekonomisk politik
Hjalmarsson, Lennart (1975), Studies in a Dynamic Theory of Production and its Applications
Örtendahl, Per-Anders (1975), Substitutionsaspekter på produktionsprocessen vid massaframställning
Anderson, Arne M. (1976), Produktion, kapacitet och kostnader vid ett helautomatiskt emballageglasbruk
Ohlsson, Olle (1976), Substitution och odelbarheter i produktionsprocessen vid massaframställning
Gunnarsson, Jan (1976), Produktionssystem och tätortshierarki – om sambandet mellan rumslig och ekonomisk struktur
Köstner, Evert (1976), Optimal allokering av tid mellan utbildning och arbete
Wigren, Rune (1976), Analys av regionala effektivitetsskillnader inom industribranscher
Wästlund, Jan (1976), Skattning och analys av regionala effektivitetsskillnader inom industribranscher
Flöjstad, Gunnar (1976), Studies in Distortions, Trade and Allocation Problems
Sandelin, Bo (1977), Prisutveckling och kapitalvinster på bostadsfastigheter
Dahlberg, Lars (1977), Empirical Studies in Public Planning
Lönnroth, Johan (1977), Marxism som matematisk ekonomi
Johansson, Börje (1978), Contributions to Sequential Analysis of Oligopolistic Competition

<u>Ekonomiska Studier</u>, utgivna av Nationalekonomiska institutionen vid Göteborgs Universitet. Nr 1 och 4 var inte doktorsavhandlingar. (The contributions to the department series 'Ekonomiska Studier' where no. 1 and 4 were no doctoral theses):

2. **Ambjörn, Erik** (1959), Svenskt importberoende 1926-1956: en ekonomisk-statistisk kartläggning med kommentarer
3. **Landgren, K-G.** (1960), Den "Nya ekonomien" i Sverige: J.M. Keynes, E. Wigfors och utecklingen 1927-39
5. **Bigsten, Arne** (1979), Regional Inequality and Development: A Case Study of Kenya
6. **Andersson, Lars** (1979), Statens styrning av de kommunala budgetarnas struktur (Central Government Influence on the Structure of the Municipal Budget)
7. **Gustafsson, Björn** (1979), Inkomst- och uppväxtförhållanden (Income and Family Background)

8. **Granholm, Arne** (1981), Interregional Planning Models for the Allocation of Private and Public Investments
9. **Lundborg, Per** (1982), Trade Policy and Development: Income Distributional Effects in the Less Developed Countries of the US and EEC Policies for Agricultural Commodities
10. **Juås, Birgitta** (1982), Värdering av risken för personskador. En jämförande studie av implicita och explicita värden. (Valuation of Personal Injuries. A comparison of Explicit and Implicit Values)
11. **Bergendahl, Per-Anders** (1982), Energi och ekonomi - tillämpningar av input-output analys (Energy and the Economy - Applications of Input-Output Analysis)
12. **Blomström, Magnus** (1983), Foreign Investment, Technical Efficiency and Structural Change - Evidence from the Mexican Manufacturing Industry
13. **Larsson, Lars-Göran** (1983), Comparative Statics on the Basis of Optimization Methods
14. **Persson, Håkan** (1983), Theory and Applications of Multisectoral Growth Models
15. **Sterner, Thomas** (1986), Energy Use in Mexican Industry.
16. **Flood, Lennart** (1986), On the Application of Time Use and Expenditure Allocation Models.
17. **Schuller, Bernd-Joachim** (1986), Ekonomi och kriminalitet - en empirisk undersökning av brottsligheten i Sverige (Economics of crime - an empirical analysis of crime in Sweden)
18. **Walfridson, Bo** (1987), Dynamic Models of Factor Demand. An Application to Swedish Industry.
19. **Stålhammar, Nils-Olov** (1987), Strukturomvandling, företagsbeteende och förväntningsbildning inom den svenska tillverkningsindustrin (Structural Change, Firm Behaviour and Expectation Formation in Swedish Manufactury)
20. **Anxo, Dominique** (1988), Sysselsättningseffekter av en allmän arbetstidsför-kortning (Employment effects of a general shortage of the working time)
21. **Mbelle, Ammon** (1988), Foreign Exchange and Industrial Development: A Study of Tanzania.
22. **Ongaro, Wilfred** (1988), Adoption of New Farming Technology: A Case Study of Maize Production in Western Kenya.
23. **Zejan, Mario** (1988), Studies in the Behavior of Swedish Multinationals.
24. **Görling, Anders** (1988), Ekonomisk tillväxt och miljö. Förorenings-struktur och ekonomiska effekter av olika miljövårdsprogram. (Economic Growth and Environment. Pollution Structure and Economic Effects of Some Environmental Programs).
25. **Aguilar, Renato** (1988), Efficiency in Production: Theory and an Application on Kenyan Smallholders.
26. **Kayizzi-Mugerwa, Steve** (1988), External Shocks and Adjustment in Zambia.

27. **Bornmalm-Jardelöw, Gunilla** (1988), Högre utbildning och arbetsmarknad (Higher Education and the Labour Market)
28. **Tansini, Ruben** (1989), Technology Transfer: Dairy Industries in Sweden and Uruguay.
29. **Andersson, Irene** (1989), Familjebeskattning, konsumtion och arbetsutbud - En ekonometrisk analys av löne- och inkomstelasticiteter samt policysimuleringar för svenska hushåll (Family Taxation, Consumption and Labour Supply - An Econometric Analysis of Wage and Income Elasticities and Policy Simulations for Swedish Households)
30. **Henrekson, Magnus** (1990), An Economic Analysis of Swedish Government Expenditure
31. **Sjöö, Boo** (1990), Monetary Policy in a Continuous Time Dynamic Model for Sweden
32. **Rosén, Åsa** (1991), Contributions to the Theory of Labour Contracts.
33. **Loureiro, Joao M. de Matos** (1992), Foreign Exchange Intervention, Sterilization and Credibility in the EMS: An Empirical Study
34. **Irandoust, Manuchehr** (1993), Essays on the Behavior and Performance of the Car Industry
35. **Tasiran, Ali Cevat** (1993), Wage and Income Effects on the Timing and Spacing of Births in Sweden and the United States
36. **Milopoulos, Christos** (1993), Investment Behaviour under Uncertainty: An Econometric Analysis of Swedish Panel Data
37. **Andersson, Per-Åke** (1993), Labour Market Structure in a Controlled Economy: The Case of Zambia
38. **Storrie, Donald W.** (1993), The Anatomy of a Large Swedish Plant Closure
39. **Semboja, Haji Hatibu Haji** (1993), Energy and Development in Kenya
40. **Makonnen, Negatu** (1993), Labor Supply and the Distribution of Economic Well-Being: A Case Study of Lesotho
41. **Julin, Eva** (1993), Structural Change in Rural Kenya
42. **Durevall, Dick** (1993), Essays on Chronic Inflation: The Brazilian Experience
43. **Veiderpass, Ann** (1993), Swedish Retail Electricity Distribution: A Non-Parametric Approach to Efficiency and Productivity Change
44. **Odeck, James** (1993), Measuring Productivity Growth and Efficiency with Data Envelopment Analysis: An Application on the Norwegian Road Sector
45. **Mwenda, Abraham** (1993), Credit Rationing and Investment Behaviour under Market Imperfections: Evidence from Commercial Agriculture in Zambia
46. **Mlambo, Kupukile** (1993), Total Factor Productivity Growth: An Empirical Analysis of Zimbabwe's Manufacturing Sector Based on Factor Demand Modelling
47. **Ndung'u, Njuguna** (1993), Dynamics of the Inflationary Process in Kenya
48. **Modén, Karl-Markus** (1993), Tax Incentives of Corporate Mergers and Foreign Direct Investments
49. **Franzén, Mikael** (1994), Gasoline Demand - A Comparison of Models
50. **Heshmati, Almas** (1994), Estimating Technical Efficiency, Productivity Growth And Selectivity Bias Using Rotating Panel Data: An Application to Swedish Agriculture

51. **Salas, Osvaldo** (1994), Efficiency and Productivity Change: A Micro Data Case Study of the Colombian Cement Industry
52. **Bjurek, Hans** (1994), Essays on Efficiency and Productivity Change with Applications to Public Service Production
53. **Cabezas Vega, Luis** (1994), Factor Substitution, Capacity Utilization and Total Factor Productivity Growth in the Peruvian Manufacturing Industry
54. **Katz, Katarina** (1994), Gender Differentiation and Discrimination. A Study of Soviet Wages
55. **Asal, Maher** (1995), Real Exchange Rate Determination and the Adjustment Process: An Empirical Study in the Cases of Sweden and Egypt
56. **Kjulin, Urban** (1995), Economic Perspectives on Child Care
57. **Andersson, Göran** (1995), Volatility Forecasting and Efficiency of the Swedish Call Options Market
58. **Forteza, Alvaro** (1996), Credibility, Inflation and Incentive Distortions in the Welfare State
59. **Locking, Håkan** (1996), Essays on Swedish Wage Formation
60. **Välilä, Timo** (1996), Essays on the Credibility of Central Bank Independence
61. **Yilma, Mulugeta** (1996), Measuring Smallholder Efficiency: Ugandan Coffee and Food-Crop Production
62. **Mabugu, Ramos E.** (1996), Tax Policy Analysis in Zimbabwe Applying General Equilibrium Models
63. **Johansson, Olof** (1996), Welfare, Externalities, and Taxation; Theory and Some Road Transport Applications.
64. **Chitiga, Margaret** (1996), Computable General Equilibrium Analysis of Income Distribution Policies in Zimbabwe
65. **Leander, Per** (1996), Foreign Exchange Market Behavior Expectations and Chaos
66. **Hansen, Jörgen** (1997), Essays on Earnings and Labor Supply
67. **Cotfas, Mihai** (1997), Essays on Productivity and Efficiency in the Romanian Cement Industry
68. **Horgby, Per-Johan** (1997), Essays on Sharing, Management and Evaluation of Health Risks
69. **Nafar, Nosratollah** (1997), Efficiency and Productivity in Iranian Manufacturing Industries
70. **Zheng, Jinghai** (1997), Essays on Industrial Structure, Technical Change, Employment Adjustment, and Technical Efficiency
71. **Isaksson, Anders** (1997), Essays on Financial Liberalisation in Developing Countries: Capital mobility, price stability, and savings
72. **Gerdin, Anders** (1997), On Productivity and Growth in Kenya, 1964-94
73. **Sharifi, Alimorad** (1998), The Electricity Supply Industry in Iran: Organization, performance and future development
74. **Zamanian, Max** (1997), Methods for Mutual Fund Portfolio Evaluation: An application to the Swedish market
75. **Manda, Damiano Kulundu** (1997), Labour Supply, Returns to Education, and the Effect of Firm Size on Wages: The case of Kenya
76. **Holmén, Martin** (1998), Essays on Corporate Acquisitions and Stock Market Introductions

77. **Pan, Kelvin** (1998), Essays on Enforcement in Money and Banking
78. **Rogat, Jorge** (1998), The Value of Improved Air Quality in Santiago de Chile
79. **Peterson, Stefan** (1998), Essays on Large Shareholders and Corporate Control
80. **Belhaj, Mohammed** (1998), Energy, Transportation and Urban Environment in Africa: The Case of Rabat-Salé, Morocco
81. **Mekonnen, Alemu** (1998), Rural Energy and Afforestation: Case Studies from Ethiopia
82. **Johansson, Anders** (1998), Empirical Essays on Financial and Real Investment Behavior
83. **Köhlin, Gunnar** (1998), The Value of Social Forestry in Orissa, India
84. **Levin, Jörgen** (1998), Structural Adjustment and Poverty: The Case of Kenya
85. **Ncube, Mkhululi** (1998), Analysis of Employment Behaviour in Zimbabwe
86. **Mwansa, Ladslous** (1998), Determinants of Inflation in Zambia
87. **Agnarsson, Sveinn** (1998), Of Men and Machines: Essays in Applied Labour and Production Economics
88. **Kadenge, Phineas** (1998), Essays on Macroeconomic Adjustment in Zimbabwe: Inflation, Money Demand, and the Real Exchange Rate
89. **Nyman, Håkan** (1998), An Economic Analysis of Lone Motherhood in Sweden
90. **Carlsson, Fredrik** (1999), Essays on Externalities and Transport
91. **Johansson, Mats** (1999), Empirical Studies of Income Distribution
92. **Alemu, Tekie** (1999), Land Tenure and Soil Conservation: Evidence from Ethiopia
93. **Lundvall, Karl** (1999), Essays on Manufacturing Production in a Developing Economy: Kenya 1992-94
94. **Zhang, Jianhua** (1999), Essays on Emerging Market Finance
95. **Mlima, Aziz Ponary** (1999), Four Essays on Efficiency and Productivity in Swedish Banking
96. **Davidsen, Björn-Ivar** (2000), Bidrag til den økonomisk-metodologiske tenkningen (Contributions to the Economic Methodological Thinking)
97. **Ericson, Peter** (2000), Essays on Labor Supply
98. **Söderbom, Måns** (2000), Investment in African Manufacturing: A Microeconomic Analysis
99. **Höglund, Lena** (2000), Essays on Environmental Regulation with Applications to Sweden
100. **Olsson, Ola** (2000), Perspectives on Knowledge and Growth
101. **Meuller, Lars** (2000), Essays on Money and Credit
102. **Österberg, Torun** (2000), Economic Perspectives on Immigrants and Intergenerational Transmissions
103. **Kalinda Mkenda, Beatrice** (2001), Essays on Purchasing Power Parity, RealExchange Rate, and Optimum Currency Areas
104. **Nerhagen, Lena** (2001), Travel Demand and Value of Time - Towards an Understanding of Individuals Choice Behavior

105. **Mkenda, Adolf** (2001), Fishery Resources and Welfare in Rural Zanzibar
106. **Eggert, Håkan** (2001), Essays on Fisheries Economics
107. **Andrén, Daniela** (2001), Work, Sickness, Earnings, and Early Exits from the Labor Market. An Empirical Analysis Using Swedish Longitudinal Data
108. **Nivorozhkin, Eugene** (2001), Essays on Capital Structure
109. **Hammar, Henrik** (2001), Essays on Policy Instruments: Applications to Smoking and the Environment
110. **Nannyonjo, Justine** (2002), Financial Sector Reforms in Uganda (1990-2000): Interest Rate Spreads, Market Structure, Bank Performance and Monetary Policy
111. **Wu, Hong** (2002), Essays on Insurance Economics
112. **Linde-Rahr, Martin** (2002), Household Economics of Agriculture and Forestry in Rural Vienam
113. **Maneschiöld, Per-Ola** (2002), Essays on Exchange Rates and Central Bank Credibility
114. **Andrén, Thomas** (2002), Essays on Training, Welfare and Labor Supply
115. **Granér, Mats** (2002), Essays on Trade and Productivity: Case Studies of Manufacturing in Chile and Kenya
116. **Jaldell, Henrik** (2002), Essays on the Performance of Fire and Rescue Services
117. **Alpizar, Francisco, R.** (2002), Essays on Environmental Policy-Making in Developing Countries: Applications to Costa Rica
118. **Wahlberg, Roger** (2002), Essays on Discrimination, Welfare and Labor Supply
119. **Piculescu, Violeta** (2002), Studies on the Post-Communist Transition
120. **Pylkkänen, Elina** (2003), Studies on Household Labor Supply and Home Production
121. **Löfgren, Åsa** (2003), Environmental Taxation – Empirical and Theoretical Applications
122. **Ivaschenko, Oleksiy** (2003), Essays on Poverty, Income Inequality and Health in Transition Economies
123. **Lundström, Susanna** (2003), On Institutions, Economic Growth and the Environment
124. **Wambugu, Anthony** (2003), Essays on Earnings and Human Capital in Kenya
125. **Adler, Johan** (2003), Aspects of Macroeconomic Saving
126. **Erlandsson, Mattias** (2003), On Monetary Integration and Macroeconomic Policy
127. **Brink, Anna** (2003), On the Political Economy of Municipality Break-Ups
128. **Ljungwall, Christer** (2003), Essays on China's Economic Performance During the Reform Period
129. **Chifamba, Ronald** (2003), Analysis of Mining Investments in Zimbabwe
130. **Muchapondwa, Edwin** (2003), The Economics of Community-Based Wildlife Conservation in Zimbabwe
131. **Hammes, Klaus** (2003), Essays on Capital Structure and Trade Financing
132. **Abou-Ali, Hala** (2003), Water and Health in Egypt: An Empirical Analysis
133. **Simatele, Munacinga** (2004), Financial Sector Reforms and Monetary Policy in Zambia
134. **Tezic, Kerem** (2004), Essays on Immigrants' Economic Integration
135. INSTÄLLD

136. **Gjirja, Matilda** (2004), Efficiency and Productivity in Swedish Banking
137. **Andersson, Jessica** (2004), Welfare Environment and Tourism in Developing Countries
138. **Chen, Yinghong** (2004), Essays on Voting Power, Corporate Governance and Capital Structure
139. **Yesuf, Mahmud** (2004), Risk, Time and Land Management under Market Imperfections: Applications to Ethiopia
140. **Kateregga, Eseza** (2005), Essays on the Infestation of Lake Victoria by the Water Hyacinth
141. **Edvardsen, Dag Fjeld** (2004), Four Essays on the Measurement of Productive Efficiency
142. **Lidén, Erik** (2005), Essays on Information and Conflicts of Interest in Stock Recommendations
143. **Dieden, Sten** (2005), Income Generation in the African and Coloured Population – Three Essays on the Origins of Household Incomes in South Africa
144. **Eliasson, Marcus** (2005), Individual and Family Consequences of Involuntary Job Loss
145. **Mahmud, Minhaj** (2005), Measuring Trust and the Value of Statistical Lives: Evidence from Bangladesh
146. **Lokina, Razack Bakari** (2005), Efficiency, Risk and Regulation Compliance: Applications to Lake Victoria Fisheries in Tanzania
147. **Jussila Hammes, Johanna** (2005), Essays on the Political Economy of Land Use Change
148. **Nyangena, Wilfred** (2006), Essays on Soil Conservation, Social Capital and Technology Adoption
149. **Nivorozhkin, Anton** (2006), Essays on Unemployment Duration and Programme Evaluation
150. **Sandén, Klas** (2006), Essays on the Skill Premium
151. **Deng, Daniel** (2006), Three Essays on Electricity Spot and Financial Derivative Prices at the Nordic Power Exchange
152. **Gebreeyesus, Mulu** (2006), Essays on Firm Turnover, Growth, and Investment Behavior in Ethiopian Manufacturing
153. **Islam, Nizamul Md.** (2006), Essays on Labor Supply and Poverty: A Microeconometric Application
154. **Kjaer, Mats** (2006), Pricing of Some Path-Dependent Options on Equities and Commodities
155. **Shimeles, Abebe** (2006), Essays on Poverty, Risk and Consumption Dynamics in Ethiopia
156. **Larsson, Jan** (2006), Four Essays on Technology, Productivity and Environment
157. **Congdon Fors, Heather** (2006), Essays in Institutional and Development Economics
158. **Akpalu, Wisdom** (2006), Essays on Economics of Natural Resource Management and Experiments
159. **Daruvala, Dinky** (2006), Experimental Studies on Risk, Inequality and Relative Standing
160. **García, Jorge** (2007), Essays on Asymmetric Information and Environmental Regulation through Disclosure

161. **Bezabih, Mintewab** (2007), Essays on Land Lease Markets, Productivity, Biodiversity, and Environmental Variability
162. **Visser, Martine** (2007), Fairness, Reciprocity and Inequality: Experimental Evidence from South Africa
163. **Holm, Louise** (2007), A Non-Stationary Perspective on the European and Swedish Business Cycle
164. **Herbertsson, Alexander** (2007), Pricing Portfolio Credit Derivatives
165. **Johansson, Anders C.** (2007), Essays in Empirical Finance: Volatility, Interdependencies, and Risk in Emerging Markets
166. **Ibáñez Díaz, Marcela** (2007), Social Dilemmas: The Role of Incentives, Norms and Institutions
167. **Ekbom, Anders** (2007), Economic Analysis of Soil Capital, Land Use and Agricultural Production in Kenya
168. **Sjöberg, Pål** (2007), Essays on Performance and Growth in Swedish Banking
169. **Palma Aguirre, Grisha Alexis** (2008), Explaining Earnings and Income Inequality in Chile
170. **Akay, Alpaslan** (2008), Essays on Microeconometrics and Immigrant Assimilation
171. **Carlsson, Evert** (2008), After Work – Investing for Retirement
172. **Munshi, Farzana** (2008), Essays on Globalization and Occupational Wages
173. **Tsakas, Elias** (2008), Essays on Epistemology and Evolutionary Game Theory
174. **Erlandzon, Karl** (2008), Retirement Planning: Portfolio Choice for Long-Term Investors
175. **Lampi, Elina** (2008), Individual Preferences, Choices, and Risk Perceptions – Survey Based Evidence
176. **Mitrut, Andreea** (2008), Four Essays on Interhousehold Transfers and Institutions in Post-Communist Romania
177. **Hansson, Gustav** (2008), Essays on Social Distance, Institutions, and Economic Growth
178. **Zikhali, Precious** (2008), Land Reform, Trust and Natural Resource Management in Africa
179. **Tengstam, Sven** (2008), Essays on Smallholder Diversification, Industry Location, Debt Relief, and Disability and Utility
180. **Boman, Anders** (2009), Geographic Labour Mobility – Causes and Consequences
181. **Qin, Ping** (2009), Risk, Relative Standing and Property Rights: Rural Household Decision-Making in China
182. **Wei, Jiegen** (2009), Essays in Climate Change and Forest Management
183. **Belu, Constantin** (2009), Essays on Efficiency Measurement and Corporate Social Responsibility
184. **Ahlerup, Pelle** (2009), Essays on Conflict, Institutions, and Ethnic Diversity
185. **Quiroga, Miguel** (2009), Microeconomic Policy for Development: Essays on Trade and Environment, Poverty and Education
186. **Zerfu, Daniel** (2010), Essays on Institutions and Economic Outcomes
187. **Wollbrant, Conny** (2010), Self-Control and Altruism
188. **Villegas Palacio, Clara** (2010), Formal and Informal Regulations: Enforcement and Compliance

189. **Maican, Florin** (2010), Essays in Industry Dynamics on Imperfectly Competitive Markets
190. **Jakobsson, Niklas** (2010), Laws, Attitudes and Public Policy
191. **Mănescu, Cristiana** (2010), Economic Implications of Corporate Social Responsibility and Responsible Investments
192. **He, Haoran** (2010), Environmental and Behavioral Economics – Applications to China